STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 118242

TO: David Lukton
Location: REM-3B75/3C70
Art Unit: 1653
Wednesday, March 31, 2004

Case Serial Number: 09/909077

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-B55
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Lukton,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
CM-1, Rm. 6-A-06
605-1155

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:42:17 ; Search time 14.1333 Seconds
(without alignments)
29.222 Million cell updates/sec

Title: US-09-909-077-1
Perfect score: 32
Sequence: 1 DTEDVVXX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	93.8	7	4	US-09-777-785A-1	Sequence 1, Appli
2	30	93.8	8	3	US-09-288-391-3	Sequence 3, Appli
3	30	93.8	8	3	US-09-288-391-5	Sequence 5, Appli
4	30	93.8	8	3	US-09-288-391-6	Sequence 6, Appli
5	30	93.8	8	3	US-09-288-391-8	Sequence 8, Appli
6	30	93.8	8	3	US-09-288-391-10	Sequence 10, Appl
7	30	93.8	8	3	US-09-288-391-11	Sequence 11, Appl
8	30	93.8	8	3	US-09-288-391-12	Sequence 12, Appl
9	30	93.8	8	3	US-09-288-391-13	Sequence 13, Appl
10	30	93.8	8	3	US-09-288-391-14	Sequence 14, Appl
11	30	93.8	10	3	US-09-288-391-1	Sequence 1, Appli
12	30	93.8	10	3	US-09-288-391-4	Sequence 4, Appli
13	30	93.8	10	3	US-09-288-391-9	Sequence 9, Appli
14	30	93.8	10	3	US-09-288-391-15	Sequence 15, Appl
15	30	93.8	12	3	US-09-288-391-2	Sequence 2, Appli
16	30	93.8	12	3	US-09-288-391-7	Sequence 7, Appli
17	30	93.8	13	3	US-09-288-391-16	Sequence 16, Appl
18	30	93.8	13	3	US-09-288-391-18	Sequence 18, Appl
19	30	93.8	13	3	US-09-288-391-19	Sequence 19, Appl
20	30	93.8	14	3	US-09-288-391-17	Sequence 17, Appl
21	30	93.8	14	4	US-09-344-456-3	Sequence 3, Appli
22	30	93.8	16	1	US-08-439-747A-17	Sequence 17, Appl
23	30	93.8	16	1	US-08-439-747A-31	Sequence 31, Appl
24	30	93.8	16	2	US-08-440-409B-17	Sequence 17, Appl
25	30	93.8	16	2	US-08-853-623D-25	Sequence 25, Appl
26	30	93.8	16	3	US-09-198-723A-25	Sequence 25, Appl
27	30	93.8	16	4	US-09-684-881-25	Sequence 25, Appl
28	30	93.8	17	1	US-08-571-643A-5	Sequence 5, Appli
29	30	93.8	17	1	US-08-439-747A-16	Sequence 16, Appl
30	30	93.8	17	1	US-08-439-747A-18	Sequence 18, Appl
31	30	93.8	17	2	US-08-440-409B-16	Sequence 16, Appl
32	30	93.8	17	2	US-08-440-409B-18	Sequence 18, Appl
33	30	93.8	17	2	US-08-853-623D-22	Sequence 22, Appl
34	30	93.8	17	3	US-09-198-723A-91	Sequence 91, Appl
35	30	93.8	17	3	US-09-288-391-20	Sequence 20, Appl
36	30	93.8	17	3	US-09-288-391-21	Sequence 21, Appl
37	30	93.8	17	3	US-09-288-391-24	Sequence 24, Appl
38	30	93.8	17	4	US-09-684-881-91	Sequence 91, Appl
39	30	93.8	18	1	US-08-439-747A-19	Sequence 19, Appl
40	30	93.8	18	1	US-08-439-747A-20	Sequence 20, Appl
41	30	93.8	18	2	US-08-440-409B-19	Sequence 19, Appl
42	30	93.8	18	2	US-08-440-409B-20	Sequence 20, Appl
43	30	93.8	18	3	US-09-144-759-10	Sequence 10, Appl
44	30	93.8	18	4	US-09-570-267-10	Sequence 10, Appl
45	30	93.8	20	2	US-08-432-693-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-777-785A-1
; Sequence 1, Application US/09777785A
; Patent No. 6624290
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; TITLE OF INVENTION: Azapeptides useful in the treatment of Hepatitis C
; FILE REFERENCE: IN01130-K1 US
; CURRENT APPLICATION NUMBER: US/09/777,785A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (7)..(7)
; OTHER INFORMATION: -Pro-NH-N(Pr.)-(C=O)-O-Ph.-(4-NO2)
US-09-777-785A-1

Query Match 93.8%; Score 30; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 2
US-09-288-391-3
; Sequence 3, Application US/09288391
; Patent No. 6251583
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; APPLICANT: Malcolm, Bruce
; APPLICANT: Beyer, Brian
; APPLICANT: Njoroge, George
; APPLICANT: Durkin, James
; APPLICANT: Windsor, William
; TITLE OF INVENTION: No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth

```
;       STATE:  New Jersey
;       COUNTRY:  USA
;       ZIP:  07033
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/288,391
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McLaughlin, Jaye P.
;       REGISTRATION NUMBER:  41,211
;       REFERENCE/DOCKET NUMBER:  IN0829P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908)298-5056
;       TELEFAX:  (908)298-5388
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  /note= The aspartic acid residue at position 1 is N-acety
US-09-288-391-3

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 3
US-09-288-391-5
; Sequence 5, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Abu (a-aminobutyric acid).  T
US-09-288-391-5

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 4
US-09-288-391-6
; Sequence 6, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-6

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 5
US-09-288-391-8
; Sequence 8, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-8

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 6
US-09-288-391-10
; Sequence 10, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-10

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 7
US-09-288-391-11
; Sequence 11, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-11

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 8
US-09-288-391-12
; Sequence 12, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-12

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 9
US-09-288-391-13
; Sequence 13, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-13

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 10
US-09-288-391-14
; Sequence 14, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
```

```
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The aspartic acid residue at position 1 is N-acety
US-09-288-391-14

  Query Match             93.8%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 11
US-09-288-391-1
; Sequence 1, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The cysteine residue at position 10 is modified as
US-09-288-391-1

  Query Match             93.8%;  Score 30;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 12
US-09-288-391-4
; Sequence 4, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
```

```
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  /note= The glycine residue at position 1 is N-acetylated.
US-09-288-391-4

  Query Match             93.8%;  Score 30;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 13
US-09-288-391-9
; Sequence 9, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 10 is Nva (norvaline).  The Nva re
US-09-288-391-9

  Query Match             93.8%;  Score 30;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 14
US-09-288-391-15
; Sequence 15, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 10 is Nva (norvaline). The Nva res
US-09-288-391-15

  Query Match             93.8%;  Score 30;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 15
US-09-288-391-2
; Sequence 2, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
```

; ZIP: 07033
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/288,391
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: IN0829P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note= The cysteine residue at position 12 is modified as
US-09-288-391-2

Query Match 93.8%; Score 30; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 5 DTEDVV 10

Search completed: March 31, 2004, 16:50:32
Job time : 14.1333 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:40:57 ; Search time 34.1333 Seconds
(without alignments)
73.950 Million cell updates/sec

Title: US-09-909-077-3
Perfect score: 35
Sequence: 1 DTEDVVAX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	97.1	344	16	Q8PDF0	Q8pdf0 xanthomonas
2	33	94.3	317	16	Q9AK29	Q9ak29 streptomyce
3	33	94.3	317	16	Q82HZ8	Q82hz8 streptomyce
4	31	88.6	235	16	Q8NQ61	Q8nq61 corynebacte
5	31	88.6	439	5	O44918	O44918 caenorhabdi
6	31	88.6	464	16	Q8PCL6	Q8pcl6 xanthomonas
7	30	85.7	181	12	Q81730	Q81730 hepatitis c
8	30	85.7	235	11	Q8BSU3	Q8bsu3 mus musculu
9	30	85.7	372	12	Q8V4W9	Q8v4w9 monkeypox v
10	30	85.7	372	12	Q8JLA6	Q8jla6 ectromelia
11	30	85.7	372	12	O57219	O57219 vaccinia vi
12	30	85.7	372	12	Q8QMU4	Q8qmu4 cowpox viru
13	30	85.7	372	12	Q80HW0	Q80hw0 vaccinia vi
14	30	85.7	372	12	Q80DW5	Q80dw5 cowpox viru
15	30	85.7	372	12	Q8V2P9	Q8v2p9 camelpox vi
16	30	85.7	380	16	Q8CW50	Q8cw50 escherichia
17	30	85.7	410	5	Q9VI96	Q9vi96 drosophila
18	30	85.7	412	5	O01819	O01819 caenorhabdi
19	30	85.7	424	6	Q864C2	Q864c2 macaca fasc
20	30	85.7	448	12	Q9E3A2	Q9e3a2 hepatitis c
21	30	85.7	448	12	Q9E364	Q9e364 hepatitis c
22	30	85.7	448	12	Q9E3D6	Q9e3d6 hepatitis c
23	30	85.7	448	12	Q9E3A6	Q9e3a6 hepatitis c
24	30	85.7	448	12	Q9E388	Q9e388 hepatitis c
25	30	85.7	448	12	Q9E363	Q9e363 hepatitis c
26	30	85.7	448	12	Q9E3D8	Q9e3d8 hepatitis c
27	30	85.7	448	12	Q9E3A7	Q9e3a7 hepatitis c
28	30	85.7	448	12	Q9E3A4	Q9e3a4 hepatitis c
29	30	85.7	448	12	Q9E3D2	Q9e3d2 hepatitis c
30	30	85.7	448	12	Q9E385	Q9e385 hepatitis c
31	30	85.7	448	12	Q9E398	Q9e398 hepatitis c
32	30	85.7	448	12	Q9E3D4	Q9e3d4 hepatitis c
33	30	85.7	448	12	Q9E381	Q9e381 hepatitis c
34	30	85.7	448	12	Q9E392	Q9e392 hepatitis c
35	30	85.7	448	12	Q9E3A8	Q9e3a8 hepatitis c
36	30	85.7	448	12	Q9E390	Q9e390 hepatitis c
37	30	85.7	448	12	Q9E3D7	Q9e3d7 hepatitis c
38	30	85.7	448	12	Q9E3A1	Q9e3a1 hepatitis c
39	30	85.7	448	12	Q9E387	Q9e387 hepatitis c
40	30	85.7	448	12	Q9E3B0	Q9e3b0 hepatitis c
41	30	85.7	448	12	Q9E3E2	Q9e3e2 hepatitis c
42	30	85.7	448	12	Q9E391	Q9e391 hepatitis c
43	30	85.7	448	12	Q9E3A9	Q9e3a9 hepatitis c
44	30	85.7	448	12	Q9E3B1	Q9e3b1 hepatitis c
45	30	85.7	448	12	Q9E394	Q9e394 hepatitis c

ALIGNMENTS

RESULT 1
Q8PDF0
ID Q8PDF0 PRELIMINARY; PRT; 344 AA.
AC Q8PDF0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Biotin synthase.
GN BIOB OR XCC0388.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012135; AAM39707.1; -.
DR GO; GO:0004076; F:biotin synthase activity; IEA.
DR GO; GO:0009102; P:biotin biosynthesis; IEA.
DR InterPro; IPR002684; Biotin_synth.

```
DR   InterPro; IPR007197; Radical_SAM.
DR   Pfam; PF04055; Radical_SAM; 1.
DR   TIGRFAMs; TIGR00433; bioB; 1.
KW   Complete proteome.
SQ   SEQUENCE   344 AA;  37515 MW;  4C576AF23C7F7D76 CRC64;

  Query Match             97.1%;  Score 34;  DB 16;  Length 344;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db         79 DTEDVVA 85


RESULT 2
Q9AK29
ID   Q9AK29      PRELIMINARY;      PRT;   317 AA.
AC   Q9AK29;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative transcriptional regulator.
GN   SCO4897 OR 2SCK8.23.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K., Harris D.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR   EMBL; AL939121; CAC33062.1; -.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR005119; LysR_subst.
DR   Pfam; PF00126; HTH_1; 1.
DR   Pfam; PF03466; LysR_substrate; 1.
DR   PRINTS; PR00039; HTHLYSR.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW   DNA-binding; Transcription regulation; Complete proteome.
SQ   SEQUENCE   317 AA;  34335 MW;  8814407AE0FD5E46 CRC64;

  Query Match             94.3%;  Score 33;  DB 16;  Length 317;
  Best Local Similarity   85.7%;  Pred. No. 59;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:||
Db         16 DTEDIVA 22


RESULT 3
Q82HZ8
ID   Q82HZ8      PRELIMINARY;      PRT;   317 AA.
AC   Q82HZ8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative LysR-family transcriptional regulator.
GN   SAV3360.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=22608306; PubMed=12692562;
RA   Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA   Sakaki Y., Hattori M., Omura S.;
RT   "Complete genome sequence and comparative analysis of the industrial
RT   microorganism Streptomyces avermitilis.";
RL   Nat. Biotechnol. 21:526-531(2003).
DR   EMBL; AP005034; BAC71071.1; -.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR005119; LysR_subst.
DR   Pfam; PF00126; HTH_1; 1.
DR   Pfam; PF03466; LysR_substrate; 1.
DR   PRINTS; PR00039; HTHLYSR.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW   Complete proteome.
SQ   SEQUENCE   317 AA;  34609 MW;  0F78A18C86596C97 CRC64;

  Query Match             94.3%;  Score 33;  DB 16;  Length 317;
  Best Local Similarity   85.7%;  Pred. No. 59;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:||
Db         16 DTEDIVA 22


RESULT 4
Q8NQ61
ID   Q8NQ61      PRELIMINARY;      PRT;   235 AA.
AC   Q8NQ61;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE   6-phosphogluconolactonase/glucosamine-6-phosphat e isomerase/deaminase
DE   (EC 3.1.1.31).
GN   CGL1578.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005278; BAB98971.1; -.
DR   GO; GO:0017057; F:6-phosphogluconolactonase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0016853; F:isomerase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   GO; GO:0006098; P:pentose-phosphate shunt; IEA.
DR   InterPro; IPR006148; Gluc_gal_isom.
DR   InterPro; IPR005900; Phosphogluconlac.
DR   Pfam; PF01182; Glucosamine_iso; 1.
DR   TIGRFAMs; TIGR01198; pgl; 1.
KW   Isomerase; Hydrolase; Complete proteome.
SQ   SEQUENCE   235 AA;  24478 MW;  8D39AE0174B6BE98 CRC64;

  Query Match             88.6%;  Score 31;  DB 16;  Length 235;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:||
Db          9 DTEDLVA 15


RESULT 5
O44918
ID   O44918      PRELIMINARY;      PRT;   439 AA.
AC   O44918;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   W10G11.17.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Goela D., Scheet P.;
RT   "The sequence of C. elegans cosmid W10G11.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF040661; AAG24223.1; -.
DR   PIR; G88103; G88103.
DR   WormPep; W10G11.17; CE14832.
KW   Hypothetical protein.
SQ   SEQUENCE   439 AA;  49291 MW;  045B32C0804DEB0E CRC64;

  Query Match             88.6%;  Score 31;  DB 5;  Length 439;
  Best Local Similarity   85.7%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||||:
Db         40 DTEDVVS 46


RESULT 6
Q8PCL6
ID   Q8PCL6      PRELIMINARY;      PRT;   464 AA.
AC   Q8PCL6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   D-lactate dehydrogenase.
GN   DLD OR XCC0708.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE012169; AAM40024.1; -.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR004113; FAD-oxidase_C.
DR   InterPro; IPR006094; Oxid_FAD_bind.
DR   Pfam; PF02913; FAD-oxidase_C; 1.
DR   Pfam; PF01565; FAD_binding_4; 1.
KW   Complete proteome.
SQ   SEQUENCE   464 AA;  47837 MW;  1FDF05C9B2492ED0 CRC64;

  Query Match             88.6%;  Score 31;  DB 16;  Length 464;
  Best Local Similarity   85.7%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||:||||
Db         51 DTDDVVA 57


RESULT 7
Q81730
ID   Q81730      PRELIMINARY;      PRT;   181 AA.
AC   Q81730;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Potential NS5 domain; putative (Genome polyprotein)
DE   (Fragment).
OS   Hepatitis C virus.
```

```
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hutchinson;
RX   MEDLINE=91013116; PubMed=2170712;
RA   Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA   Yoshizawa H.;
RT   "The 5'-terminal sequence of the hepatitis C virus genome.";
RL   Jpn. J. Exp. Med. 60:167-177(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hutchinson;
RA   Inchauspe G., Zebedee S.L, Nasoff M.S., Sugitani M., Abe K.,
RA   Prince A.M.;
RT   "Cloning and nucleotide sequence analysis of structural and
RT   nonstructural regions of the hutchinson strain of hepatitus C.";
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hutchinson;
RX   MEDLINE=89222455; PubMed=2496467;
RA   Kuo G., Choo Q.-L., Alter H.J., Gitnick G.L., Redeker A.G.,
RA   Purcell R.H., Miyamura T., Dienstag J.L., Alter M.J., Stevens C.E.,
RA   Tegtmeier G.E., Bonino F., Colombo M., Lee W.-S., Kuo C., Berger K.,
RA   Shuster J.R., Overby L.R., Bradley D.W., Houghton M.;
RT   "An assay for circulating antibodies to a major etiologic virus of
RT   human non-A, non-B hepatitis.";
RL   Science 244:362-365(1992).
DR   EMBL; M55974; AAA45663.1; -.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   InterPro; IPR002166; HCV_RdRP.
DR   Pfam; PF00998; Viral_RdRP; 1.
KW   Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW   Transferase.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19124 MW;  7C30235E19009C4A CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         98 DTEDVV 103


RESULT 8
Q8BSU3
ID   Q8BSU3      PRELIMINARY;      PRT;   235 AA.
AC   Q8BSU3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Pituitary;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK030512; BAC26998.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   235 AA;  25133 MW;  50928397A1AD126F CRC64;

  Query Match             85.7%;  Score 30;  DB 11;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         92 DTEDVV 97


RESULT 9
Q8V4W9
ID   Q8V4W9      PRELIMINARY;      PRT;   372 AA.
AC   Q8V4W9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   A7L.
GN   A7L.
OS   Monkeypox virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10244;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Zaire-96-I-16;
RX   MEDLINE=21592287; PubMed=11734207;
RA   Shchelkunov S.N., Totmenin A.V., Babkin I.V., Safronov P.F.,
RA   Ryazankina O.I., Petrov N.A., Gutorov V.V., Uvarova E.A.,
RA   Mikheev M.V., Sisler J.R., Esposito J.J., Jahrling P.B., Moss B.,
RA   Sandakhchiev L.S.;
RT   "Human monkeypox and smallpox viruses: genomic comparison.";
RL   FEBS Lett. 509:66-70(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Zaire-96-I-16;
RA   Shchelkunov S.N., Totmenin A.V., Safronov P.F., Gutorov V.V.,
RA   Ryazankina O.I., Petrov N.A., Babkin I.V., Uvarova E.A.,
RA   Esposito J.J., Moss B., Sisler J.R., Jahrling P.B., Sandakhchiev L.S.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF380138; AAL40575.1; -.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43181 MW;  6736C01472BC8583 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 10
Q8JLA6
ID   Q8JLA6      PRELIMINARY;      PRT;   372 AA.
AC   Q8JLA6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   EVM109.
GN   EVM109.
OS   Ectromelia virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=12643;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Moscow;
RX   MEDLINE=95266283; PubMed=7747448;
RA   Mossman K., Upton C., Buller R.M., McFadden G.;
RT   "Species specificity of ectromelia virus and vaccinia virus
RT   interferon-gamma binding proteins.";
RL   Virology 208:762-769(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Moscow;
RX   MEDLINE=98154919; PubMed=9495531;
RA   Wall E.M., Cao J., Chen N., Buller R.M., Upton C.;
RT   "A novel poxvirus gene and its human homolog are similar to an E. coli
RT   lysophospholipase.";
RL   Virus Res. 52:157-167(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Moscow;
RX   MEDLINE=20192152; PubMed=10725549;
RA   Chen N., Buller R.M., Wall E.M., Upton C.;
RT   "Analysis of host response modifier ORFs of ectromelia virus, the
RT   causative agent of mousepox.";
RL   Virus Res. 66:155-173(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Moscow;
RA   Chen N., Danila M.I., Feng Z., Buller M.L., Wang C., Han X.,
RA   Lefkowitz E., Upton C.;
RT   "The Genomic Sequence of Ectromelia Virus, the Causative Agent of
RT   Mousepox.";
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF012825; AAM92413.1; -.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43124 MW;  31B672F22A84DAE4 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 11
O57219
ID   O57219      PRELIMINARY;      PRT;   372 AA.
AC   O57219;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative 43.1k protein.
GN   MVA117L.
OS   Vaccinia virus (strain Ankara).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=126794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ankara;
RA   Antoine G., Scheiflinger F., Falkner F.G., Dorner F.;
RT   "The complete genomic sequence of the Modified Vaccinia Ankara (MVA)
RT   strain.";
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U94848; AAB96459.1; -.
DR   PIR; T37393; T37393.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43099 MW;  16F7DD096BB78D74 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 12
Q8QMU4
ID   Q8QMU4      PRELIMINARY;      PRT;   372 AA.
AC   Q8QMU4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   V131.
OS   Cowpox virus (CPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10243;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brighton Red;
RX   MEDLINE=83117629; PubMed=6961398;
RA   Pickup D.J., Bastia D., Stone H.O., Joklik W.K.;
RT   "Sequence of terminal regions of cowpox virus DNA: arrangement of
RT   repeated and unique sequence elements.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:7112-7116(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brighton Red;
RX   MEDLINE=90177240; PubMed=2309453;
RA   Parsons B.L., Pickup D.J.;
RT   "Transcription of orthopoxvirus telomeres at late times during
RT   infection.";
RL   Virology 175:69-80(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brighton Red;
RX   MEDLINE=91196263; PubMed=2014645;
RA   Hu F.Q., Pickup D.J.;
RT   "Transcription of the terminal loop region of vaccinia virus DNA is
RT   initiated from the telomere sequences directing DNA resolution.";
RL   Virology 181:716-720(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brighton Red;
RX   MEDLINE=94378510; PubMed=8091665;
RA   Hu F.Q., Smith C.A., Pickup D.J.;
RT   "Cowpox virus contains two copies of an early gene encoding a soluble
RT   secreted form of the type II TNF receptor.";
RL   Virology 204:343-356(1994).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brighton Red;
RA   Pickup D.J.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brighton Red;
RA   Dietrich F.S., Ray C.A., Sharma A.D., Allen A., Pickup D.J.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF482758; AAM13579.1; -.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43152 MW;  C2292597BB52B8E2 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
```

```
Db        127 DTEDIVS 133


RESULT 13
Q80HW0
ID   Q80HW0      PRELIMINARY;      PRT;   372 AA.
AC   Q80HW0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein VACWR125.
GN   VACWR125.
OS   Vaccinia virus (strain WR).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Esposito J.J., Frace A.M., Sammons S.A., Olsen-Rasmussen M.,
RA   Osborne J., Wohlhueter R.;
RT   "Sequencing of the coding region of Vaccinia-WR to an average 9-fold
RT   redundancy and an error rate of 0.16/10kb.";
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY243312; AAO89404.1; -.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   372 AA;  43174 MW;  ACBCA5186CD80264 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 14
Q80DW5
ID   Q80DW5      PRELIMINARY;      PRT;   372 AA.
AC   Q80DW5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   A7L protein.
GN   A7L.
OS   Cowpox virus (CPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10243;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=97068532; PubMed=8963248;
RA   Safronov P.F., Petrov N.A., Riazankina O.I., Totmenin A.V.,
RA   Shchelkunov S.N., Sandakhchiev L.S.;
RT   "Genes of a circle of hosts for the cowpox virus.";
RL   Dokl. Akad. Nauk 349:829-833(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=98229462; PubMed=9568042;
RA   Shchelkunov S.N., Safronov P.F., Totmenin A.V., Petrov N.A.,
RA   Ryazankina O.I., Gutorov V.V., Kotwal G.J.;
RT   "Species-specific differences in genome organization of cowpox,
RT   smallpox, and vaccinia viruses.";
RL   Virology 243:432-460(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RA   Shchelkunov S.N., Safronov P.F., Totmenin A.V., Miheev M.V.,
RA   Ryazankina O.I., Petrov N.A., Gutorov V.V., Kotwal G.J.,
RA   Sandakhchiev L.S.;
RT   "Structure-function and organization of cowpox virus strain GRI-90
RT   complete genome.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RA   Totmenin A.V.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X94355; CAD90674.1; -.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43159 MW;  F64668FDFD849831 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 15
Q8V2P9
ID   Q8V2P9      PRELIMINARY;      PRT;   372 AA.
AC   Q8V2P9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (CMP122L).
GN   CMP122L.
OS   Camelpox virus (strain CP-1).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=203174;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M-96;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Zaitsev V.L.,
RA   Kerembekova U.Z., Sandybaev N.T., Kutish G.F., Rock D.L.;
RT   "The genome of camelpox virus.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMS;
RX   PubMed=11907336;
RA   Gubser C., Smith G.L.;
RT   "The sequence of camelpox virus shows it is most closely related to
RT   variola virus, the cause of smallpox.";
RL   J. Gen. Virol. 83:855-872(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMS;
RA   Gubser C., Smith G.L.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF438165; AAL73830.1; -.
DR   EMBL; AY009089; AAG37607.1; -.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   372 AA;  43142 MW;  94DFA2E8EAAD8504 CRC64;

  Query Match             85.7%;  Score 30;  DB 12;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


Search completed: March 31, 2004, 16:48:37
```

Job time : 35.1333 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:01 ; Search time 50.6667 Seconds
(without alignments)
44.613 Million cell updates/sec

Title: US-09-909-077-4
Perfect score: 38
Sequence: 1 DTEDVVPX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	97.4	7	4	AAG66392	Aag66392 Azapeptid
2	37	97.4	7	5	AAM51807	Aam51807 HCV prote
3	37	97.4	7	6	ABR61795	Abr61795 HCV prote
4	37	97.4	8	4	AAE10054	Aae10054 Hepatitis
5	37	97.4	8	4	AAE10056	Aae10056 Hepatitis
6	37	97.4	8	5	ABB07111	Abb07111 Hepatitis
7	34	89.5	2358	6	ABU17839	Abu17839 Protein e
8	33	86.8	415	4	ABB52751	Abb52751 Escherich
9	33	86.8	548	6	ABJ25469	Abj25469 Aspergill
10	33	86.8	555	6	ABJ26069	Abj26069 Aspergill
11	33	86.8	724	6	ABR53076	Abr53076 Protein s
12	33	86.8	1204	4	AAG85003	Aag85003 Shrimp wh
13	32	84.2	95	3	AAB40526	Aab40526 Human ORF
14	31	81.6	51	5	ABP07261	Abp07261 Human ORF
15	31	81.6	96	4	AAU51779	Aau51779 Propionib
16	31	81.6	96	6	ABM48298	Abm48298 Propionib
17	31	81.6	264	3	AAG48015	Aag48015 Arabidops
18	31	81.6	264	3	AAG31756	Aag31756 Arabidops
19	31	81.6	282	4	ABG02219	Abg02219 Novel hum
20	31	81.6	290	4	ABB72011	Abb72011 Drosophil
21	31	81.6	312	3	AAG14164	Aag14164 Arabidops
22	31	81.6	312	3	AAG31755	Aag31755 Arabidops
23	31	81.6	312	3	AAG48014	Aag48014 Arabidops
24	31	81.6	384	3	AAG14163	Aag14163 Arabidops
25	31	81.6	393	3	AAG31754	Aag31754 Arabidops
26	31	81.6	393	3	AAG48013	Aag48013 Arabidops
27	31	81.6	406	3	AAG14162	Aag14162 Arabidops
28	31	81.6	423	4	AAU01860	Aau01860 Mycoplasm
29	31	81.6	423	5	AAO15862	Aao15862 Mutant My
30	31	81.6	451	4	AAU01859	Aau01859 Mycoplasm
31	31	81.6	451	5	AAO15861	Aao15861 Mycoplasm
32	31	81.6	3898	2	AAR06996	Aar06996 Protein c
33	31	81.6	3898	2	AAR60543	Aar60543 55 kiloda
34	31	81.6	3898	2	AAR95239	Aar95239 Hcg chole
35	31	81.6	3898	4	AAB20382	Aab20382 Hcg chole
36	31	81.6	4675	5	ABP70085	Abp70085 Human NOV
37	31	81.6	4691	5	ABP70084	Abp70084 Human NOV
38	30	78.9	7	5	AAM51805	Aam51805 HCV prote
39	30	78.9	7	5	AAM51806	Aam51806 HCV prote
40	30	78.9	8	4	AAE10046	Aae10046 Hepatitis
41	30	78.9	8	4	AAE10053	Aae10053 Hepatitis
42	30	78.9	8	4	AAE10055	Aae10055 Hepatitis
43	30	78.9	8	4	AAE10048	Aae10048 Hepatitis
44	30	78.9	8	4	AAE10057	Aae10057 Hepatitis
45	30	78.9	8	4	AAE10051	Aae10051 Hepatitis

ALIGNMENTS

RESULT 1
AAG66392
ID AAG66392 standard; peptide; 7 AA.
XX
AC AAG66392;
XX
DT 15-OCT-2001 (first entry)
XX
DE Azapeptide #1 useful for treating hepatitis C infection.
XX
KW Virucide; hepatotropic; azapeptide; Hepatitis C viral infection;
KW serine protease inhibitor; antiviral.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 7
FT /note= "Optionally: proline 2-[(4-nitrophenoxy)carbonyl]-
FT 2-propylhydrazine, or:proline 2-[1,2,2,2-
FT tetrachloroethoxy)- carbonyl]-2-propylhydrazine"
XX
PN WO200158929-A1.
XX
PD 16-AUG-2001.
XX
PF 06-FEB-2001; 2001WO-US003768.
XX
PR 08-FEB-2000; 2000US-0181017P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R;
XX
DR WPI; 2001-536524/59.
XX
PT New azapeptide compounds, are inhibitors of hepatitis C virus serine
PT protease, useful in treating hepatitis C infection.
XX
PS Claim 16; Page 35; 44pp; English.
XX
CC The present invention relates to novel azapeptide compounds. The present
CC sequence is one such peptide compound. This peptide compound is a
CC Hepatitis C virus serine protease inhibitor. This peptide compound can be
CC used in the treatment of hepatitis C viral infections, by administrating
CC the present peptide compound, optionally in combination with an antiviral
CC agent and/or an enzyme inhibitor

```
XX
SQ   Sequence 7 AA;

  Query Match             97.4%;  Score 37;  DB 4;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 2
AAM51807
ID   AAM51807 standard; peptide; 7 AA.
XX
AC   AAM51807;
XX
DT   22-JAN-2002  (first entry)
XX
DE   HCV protease inhibition assay substrate peptide #3.
XX
KW   HCV; Hepatitis C virus; virucide; hepatotropic; antiinflammatory;
KW   Hepatitis C; NS3/NS4a serine protease.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /label= OTHER
FT                   /note= "N-terminal acetyl"
FT   Modified-site   7
FT                   /label= OTHER
FT                   /note= "modified by Nva"
XX
PN   WO200177113-A2.
XX
PD   18-OCT-2001.
XX
PF   03-APR-2001; 2001WO-US010869.
XX
PR   05-APR-2000; 2000US-0194607P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Chen KX,  Arasappan A,  Venkatraman S,  Parekh TN,  Gu H,  Njoroge FG;
PI   Girijavallabhan VM,  Ganguly A,  Saksena A,  Jao E,  Yao NH,  Prongay AJ;
PI   Madison VS,  Vibulbhan B;
XX
DR   WPI; 2002-017438/02.
XX
PT   New macrocyclic compounds are hepatitis C virus inhibitors (HCV),
PT   especially HCV NS3/NS4a serine protease inhibitors, useful for treating
PT   hepatitis C and related disorders.
XX
PS   Example 111; Page 359; 402pp; English.
XX
CC   The present invention relates to macrocyclic compounds and their
CC   derivatives, which are capable of acting as Hepatitis C virus (HCV)
CC   inhibitors. They are particularly useful for inhibiting HCV NS3/NS4a
CC   serine protease. The compounds can be used to treat disorders associated
CC   with HCV, including hepatitis C. The present sequence is a peptide
CC   substrate used in a HCV protease inhibition assay in the exemplification
CC   of the invention
XX
SQ   Sequence 7 AA;

  Query Match             97.4%;  Score 37;  DB 5;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 3
ABR61795
ID   ABR61795 standard; peptide; 7 AA.
XX
AC   ABR61795;
XX
DT   12-SEP-2003  (first entry)
XX
DE   HCV protein derived peptide substrate.
XX
KW   HCV; protease; drug monitoring; therapeutic; retroviral;
KW   protease inhibitor.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-terminal acetylation"
FT   Modified-site   7
FT                   /note= "Pro-(Nva)-O-4-phenylazophenyl ester"
XX
PN   WO2003040390-A2.
XX
PD   15-MAY-2003.
XX
PF   08-NOV-2002; 2002WO-EP012631.
XX
PR   08-NOV-2001; 2001US-0331117P.
XX
PA   (TIBO-) TIBOTEC PHARM LTD.
XX
PI   Gulnik S,  Yu B,  Erickson JW;
XX
DR   WPI; 2003-493269/46.
XX
PT   Determining the inhibitory potency of an active ingredient in a
PT   biological sample, useful for therapeutic drug monitoring comprises
PT   relating the signal determined to a reference standard curve prepared
PT   with at least one reference.
XX
PS   Disclosure; Page 7; 32pp; English.
XX
CC   The invention relates to determining the inhibitory potency of an active
CC   ingredient in a biological sample. The method involves providing a
CC   biological sample, a bioactive molecule and a reagent for the bioactive
CC   molecule, which are then added to a container, determining the signal and
CC   relating the signal to a reference standard curve prepared with at least
CC   one reference. The methods are useful for therapeutic drug monitoring by
CC   determining the amount or concentration of protease inhibitors, including
CC   retroviral protease inhibitors such as HIV inhibitors. The present
CC   sequence represents a substrate containing a peptide backbone derived
CC   from HCV proteins
XX
SQ   Sequence 7 AA;

  Query Match             97.4%;  Score 37;  DB 6;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 4
AAE10054
ID   AAE10054 standard; peptide; 8 AA.
XX
AC   AAE10054;
XX
```

```
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #11.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   8
FT                   /label= Nva
FT                   /note= "Nva-O-7-hydroxy-4-methyl-coumarin"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             97.4%;  Score 37;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 5
AAE10056
ID   AAE10056 standard; peptide; 8 AA.
XX
AC   AAE10056;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #13.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   8
FT                   /label= Nva
FT                   /note= "Nva-O-4-phenylazophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             97.4%;  Score 37;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 6
ABB07111
ID   ABB07111 standard; peptide; 8 AA.
XX
AC   ABB07111;
XX
DT   25-JUN-2002  (first entry)
XX
DE   Hepatitis C virus NS3-serine protease inhibitor related peptide #35.
XX
KW   Hepatitis C virus; HCV; NS3-serine protease inhibitor; hepatitis;
KW   HCV protease inhibitor; infection; virucide; hepatotrophic.
```

XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "acetylated"
FT Modified-site 8
FT /label= Nva
FT /note= "norvaline"
XX
PN WO200208256-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US022826.
XX
PR 21-JUL-2000; 2000US-0220109P.
XX
PA (SCHE) SCHERING CORP.
PA (CORV-) CORVAS INT INC.
XX
PI Saksena AK, Girijavallabhan VM, Lovey RG, Jao EE, Bennett F;
PI Mccormick J, Wang H, Pike RE, Bogen SL, Liu Y, Arasappan A;
PI Parekh T, Pinto PA, Njoroge FG, Ganguly AK, Brunck TK, Kemp SJ;
PI Levy OE, Lim-Wilby M;
XX
DR WPI; 2002-361644/39.
XX
PT Novel peptide inhibitor compounds of hepatitis virus NS3/NS4a serine
PT protease, useful for treating hepatitis C virus disorders.
XX
PS Example 9; Page 135; 196pp; English.
XX
CC The present invention describes a peptide compound (I) exhibiting
CC hepatitis C virus (HCV) protease inhibitory activity, including
CC enantiomers, stereoisomers, rotomers and tautomers, pharmaceutically
CC acceptable salts, solvates or derivatives. Also described are: (1) a
CC pharmaceutical composition (II) comprising (I); and (2) preparing (II)
CC for treating disorders associated with HCV protease involving bringing
CC into intimate contact (I) and a carrier. (I) has virucide and
CC hepatotrophic activities and can be used as HCV NS3/NS4a serine protease
CC inhibitors. (I) is useful for manufacturing a medicament to treat
CC disorders associated with HCV protease. (I) can be used for modulating
CC activity of HCV protease preferably, HCV NS3/NS4a protease and for
CC modulating the processing of HCV polypeptide. (II) is useful for treating
CC disorders associated with HCV and for treating disorders associated with
CC HCV protease. (I) is useful for treating hepatitis caused by HCV. The
CC present sequence represents a peptide given in an example from the
CC present invention
XX
SQ Sequence 8 AA;

Query Match 97.4%; Score 37; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
|||||||
Db 1 DTEDVVP 7

RESULT 7
ABU17839
ID ABU17839 standard; protein; 2358 AA.
XX
AC ABU17839;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #3366.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Bacillus anthracis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA21709.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 45763; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2358 AA;

Query Match 89.5%; Score 34; DB 6; Length 2358;
Best Local Similarity 85.7%; Pred. No. 7.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
||||:|
Db 92 DTEDVLP 98

RESULT 8
ABB52751

ID ABB52751 standard; protein; 415 AA.
XX
AC ABB52751;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polypeptide SEQ ID NO 903.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP003445.
XX
PR 10-MAR-2000; 2000FR-00003145.
PR 02-FEB-2001; 2001FR-00001449.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the phylogenic
PT determination of a given strain comprises polynucleotides of nature B2/D+
PT A-.
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
CC encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more frequent
CC use of broad spectrum antibiotics
XX
SQ Sequence 415 AA;

Query Match 86.8%; Score 33; DB 4; Length 415;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
||||::|
Db 79 DTEDILP 85

RESULT 9
ABJ25469
ID ABJ25469 standard; protein; 548 AA.
XX
AC ABJ25469;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #127.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 548 AA;

Query Match 86.8%; Score 33; DB 6; Length 548;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
|:|||||
Db 75 DSEDVVP 81

RESULT 10
ABJ26069
ID ABJ26069 standard; protein; 555 AA.
XX
AC ABJ26069;
XX
DT 16-APR-2003 (first entry)
XX

DE Aspergillus fumigatus essential gene protein #727.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 555 AA;

Query Match 86.8%; Score 33; DB 6; Length 555;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVP 7
              |:|||||
Db         75 DSEDVVP 81
```

RESULT 11
ABR53076
ID ABR53076 standard; protein; 724 AA.
XX
AC ABR53076;
XX
DT 20-JUN-2003 (first entry)
XX
DE Protein sequence #SEQ ID 1017.
XX
KW Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS Saccharomyces cerevisiae.
XX
PN EP1258494-A1.
XX
PD 20-NOV-2002.
XX
PF 20-DEC-2001; 2001EP-00130253.
XX
PR 15-MAY-2001; 2001EP-00111774.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;
XX
DR WPI; 2003-250078/25.
DR N-PSDB; ACC61118.
XX
PT New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or
PT disorder.
XX
PS Disclosure; SEQ ID NO 1017; 17pp + Sequence Listing; English.
XX
CC The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM
XX
SQ Sequence 724 AA;

Query Match 86.8%; Score 33; DB 6; Length 724;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVP 7
              ||:||||
Db        630 DTKDVVP 636
```

RESULT 12
AAG85003
ID AAG85003 standard; protein; 1204 AA.
XX
AC AAG85003;
XX
DT 06-AUG-2003 (revised)
DT 11-SEP-2001 (first entry)
XX
DE Shrimp white spot Bacilliform virus (WSBV) protein 94.
XX
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW antiviral agent; gene expression; antisense construct;
KW transgenic viral resistant shrimp.
XX
OS White spot syndrome virus.

XX
PN WO200138351-A2.
XX
PD 31-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US028888.
XX
PR 24-NOV-1999; 99CN-00124717.
XX
PA (PENY-) PE CORP NY.
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA (SINO-) SINOGENOMAX CO LTD.
XX
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX
DR WPI; 2001-355877/37.
DR N-PSDB; AAH62783.
XX
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (WSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating WSBV infection.
XX
PS Claim 1; Fig 3; 626pp; English.
XX
CC The invention provides the primary nucleotide sequence of the WSBV genome
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC field.)
XX
SQ Sequence 1204 AA;

Query Match 86.8%; Score 33; DB 4; Length 1204;
Best Local Similarity 71.4%; Pred. No. 5.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
||||::|
Db 743 DTEDLIP 749

RESULT 13
AAB40526
ID AAB40526 standard; protein; 95 AA.
XX
AC AAB40526;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF290 polypeptide sequence SEQ ID NO:580.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.

XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC74735.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 11; Page 721; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 95 AA;

Query Match 84.2%; Score 32; DB 3; Length 95;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
|||||:|
Db 45 DTEDVFP 51

RESULT 14
ABP07261
ID ABP07261 standard; protein; 51 AA.
XX
AC ABP07261;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:14504.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;

KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN23013.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 14504; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 51 AA;

Query Match 81.6%; Score 31; DB 5; Length 51;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
|||| ||
Db 7 DTEDPVP 13

RESULT 15
AAU51779
ID AAU51779 standard; protein; 96 AA.
XX
AC AAU51779;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #12675.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59552.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 12974; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 96 AA;

Query Match 81.6%; Score 31; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TEDVVP 7
||||||
Db 25 TEDVVP 30

Search completed: March 31, 2004, 16:45:35
Job time : 52.6667 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:42:17 ; Search time 14.1333 Seconds
(without alignments)
29.222 Million cell updates/sec

Title: US-09-909-077-4
Perfect score: 38
Sequence: 1 DTEDVVPX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	97.4	8	3	US-09-288-391-11	Sequence 11, Appl
2	37	97.4	8	3	US-09-288-391-13	Sequence 13, Appl
3	31	81.6	408	4	US-09-252-991A-31417	Sequence 31417, A
4	31	81.6	3898	2	US-08-876-991-2	Sequence 2, Appli
5	31	81.6	3898	2	US-09-059-853-2	Sequence 2, Appli
6	30	78.9	7	4	US-09-777-785A-1	Sequence 1, Appli
7	30	78.9	8	3	US-09-288-391-3	Sequence 3, Appli
8	30	78.9	8	3	US-09-288-391-5	Sequence 5, Appli
9	30	78.9	8	3	US-09-288-391-6	Sequence 6, Appli
10	30	78.9	8	3	US-09-288-391-8	Sequence 8, Appli
11	30	78.9	8	3	US-09-288-391-10	Sequence 10, Appl
12	30	78.9	8	3	US-09-288-391-12	Sequence 12, Appl
13	30	78.9	8	3	US-09-288-391-14	Sequence 14, Appl
14	30	78.9	10	3	US-09-288-391-1	Sequence 1, Appli
15	30	78.9	10	3	US-09-288-391-4	Sequence 4, Appli
16	30	78.9	10	3	US-09-288-391-9	Sequence 9, Appli
17	30	78.9	10	3	US-09-288-391-15	Sequence 15, Appl
18	30	78.9	12	3	US-09-288-391-2	Sequence 2, Appli
19	30	78.9	12	3	US-09-288-391-7	Sequence 7, Appli
20	30	78.9	13	3	US-09-288-391-16	Sequence 16, Appl
21	30	78.9	13	3	US-09-288-391-18	Sequence 18, Appl
22	30	78.9	13	3	US-09-288-391-19	Sequence 19, Appl
23	30	78.9	14	3	US-09-288-391-17	Sequence 17, Appl
24	30	78.9	14	4	US-09-344-456-3	Sequence 3, Appli
25	30	78.9	16	1	US-08-439-747A-17	Sequence 17, Appl
26	30	78.9	16	1	US-08-439-747A-31	Sequence 31, Appl
27	30	78.9	16	2	US-08-440-409B-17	Sequence 17, Appl
28	30	78.9	16	2	US-08-853-623D-25	Sequence 25, Appl
29	30	78.9	16	3	US-09-198-723A-25	Sequence 25, Appl
30	30	78.9	16	4	US-09-684-881-25	Sequence 25, Appl
31	30	78.9	17	1	US-08-571-643A-5	Sequence 5, Appli
32	30	78.9	17	1	US-08-439-747A-16	Sequence 16, Appl
33	30	78.9	17	1	US-08-439-747A-18	Sequence 18, Appl
34	30	78.9	17	2	US-08-440-409B-16	Sequence 16, Appl
35	30	78.9	17	2	US-08-440-409B-18	Sequence 18, Appl
36	30	78.9	17	2	US-08-853-623D-22	Sequence 22, Appl
37	30	78.9	17	3	US-09-198-723A-91	Sequence 91, Appl
38	30	78.9	17	3	US-09-288-391-20	Sequence 20, Appl
39	30	78.9	17	3	US-09-288-391-21	Sequence 21, Appl
40	30	78.9	17	3	US-09-288-391-24	Sequence 24, Appl
41	30	78.9	17	4	US-09-684-881-91	Sequence 91, Appl
42	30	78.9	18	1	US-08-439-747A-19	Sequence 19, Appl
43	30	78.9	18	1	US-08-439-747A-20	Sequence 20, Appl
44	30	78.9	18	2	US-08-440-409B-19	Sequence 19, Appl
45	30	78.9	18	2	US-08-440-409B-20	Sequence 20, Appl

ALIGNMENTS

```
RESULT 1
US-09-288-391-11
; Sequence 11, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-11

  Query Match             97.4%;  Score 37;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 2
US-09-288-391-13
; Sequence 13, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-13

  Query Match             97.4%;  Score 37;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 3
US-09-252-991A-31417
; Sequence 31417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 31417
;   LENGTH: 408
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31417

  Query Match             81.6%;  Score 31;  DB 4;  Length 408;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||| |||
Db        364 DTEQVVP 370


RESULT 4
US-08-876-991-2
; Sequence 2, Application US/08876991
; Patent No. 5925360
;  GENERAL INFORMATION:
;    APPLICANT:  Gregor Meyers, Tillmann R menapf,
;    APPLICANT:  Heinz-J rgen Thiel
;    TITLE OF INVENTION:  Hog cholera virus vaccine and diagnostic
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Organon Teknika Corporation
;      ADDRESSEE:  Biotechnology Research Institute
;      STREET:  l330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,991
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,577
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/650,584
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/469,702
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/123,596
;      FILING DATE:
;      APPLICATION NUMBER:  07/797,554
;      FILING DATE:  22-NOV-1991
;      APPLICATION NUMBER:  US 07/494,991
;      FILING DATE:  16-MAR-1990
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William M. Blackstone
;      REGISTRATION NUMBER:  29,772
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 258-5200
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3898 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-876-991-2
```

```
  Query Match             81.6%;  Score 31;  DB 2;  Length 3898;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||||
Db        601 TEDVVP 606


RESULT 5
US-09-059-853-2
; Sequence 2, Application US/09059853
; Patent No. 5935582
;  GENERAL INFORMATION:
;    APPLICANT:  Gregor Meyers, Tillmann R menapf,
;    APPLICANT:  Heinz-J rgen Thiel
;    TITLE OF INVENTION:  Hog cholera virus vaccine and diagnostic
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Organon Teknika Corporation
;      ADDRESSEE:  Biotechnology Research Institute
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/797,554
;      FILING DATE:  22-NOV-1991
;      APPLICATION NUMBER:  US 07/494,991
;      FILING DATE:  16-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William M. Blackstone
;      REGISTRATION NUMBER:  29,772
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 258-5200
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3898 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-059-853-2

  Query Match             81.6%;  Score 31;  DB 2;  Length 3898;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||||
Db        601 TEDVVP 606


RESULT 6
US-09-777-785A-1
; Sequence 1, Application US/09777785A
; Patent No. 6624290
; GENERAL INFORMATION:
;  APPLICANT: Zhang, Rumin
;  TITLE OF INVENTION: Azapeptides useful in the treatment of Hepatitis C
;  FILE REFERENCE: IN01130-K1 US
;  CURRENT APPLICATION NUMBER: US/09/777,785A
;  CURRENT FILING DATE:  2001-02-06
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: -Pro-NH-N(Pr.)-(C=O)-O-Ph.-(4-NO2)
US-09-777-785A-1

  Query Match             78.9%;  Score 30;  DB 4;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 7
US-09-288-391-3
; Sequence 3, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The aspartic acid residue at position 1 is N-acety
US-09-288-391-3

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
```

```
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 8
US-09-288-391-5
; Sequence 5, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Abu (a-aminobutyric acid).  T
US-09-288-391-5

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 9
US-09-288-391-6
; Sequence 6, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-6

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 10
US-09-288-391-8
; Sequence 8, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
```

```
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McLaughlin, Jaye P.
;       REGISTRATION NUMBER:  41,211
;       REFERENCE/DOCKET NUMBER:  IN0829P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908)298-5056
;       TELEFAX:  (908)298-5388
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-8

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 11
US-09-288-391-10
; Sequence 10, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-10

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 12
US-09-288-391-12
; Sequence 12, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-12

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 13
US-09-288-391-14
; Sequence 14, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The aspartic acid residue at position 1 is N-acety
US-09-288-391-14

  Query Match             78.9%;  Score 30;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 14
US-09-288-391-1
; Sequence 1, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The cysteine residue at position 10 is modified as
US-09-288-391-1

  Query Match             78.9%;  Score 30;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 15
US-09-288-391-4
; Sequence 4, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
```

```
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  /note= The glycine residue at position 1 is N-acetylated.
US-09-288-391-4

  Query Match             78.9%;  Score 30;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


Search completed: March 31, 2004, 16:50:34
Job time : 15.1333 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:45:43 ; Search time 34.4 Seconds
(without alignments)
60.852 Million cell updates/sec

Title: US-09-909-077-4
Perfect score: 38
Sequence: 1 DTEDVVPX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	97.4	7	9	US-09-777-785-1	Sequence 1, Appli
2	37	97.4	7	9	US-09-777-785-3	Sequence 3, Appli
3	37	97.4	8	10	US-09-909-062-128	Sequence 128, App
4	34	89.5	2358	12	US-10-282-122A-45763	Sequence 45763, A
5	33	86.8	72	12	US-10-424-599-197370	Sequence 197370,
6	33	86.8	372	12	US-10-425-114-47171	Sequence 47171, A
7	33	86.8	374	12	US-10-425-114-57445	Sequence 57445, A
8	33	86.8	386	12	US-10-425-114-72029	Sequence 72029, A
9	33	86.8	386	12	US-10-425-114-72218	Sequence 72218, A
10	33	86.8	415	14	US-10-238-075-903	Sequence 903, App
11	33	86.8	548	14	US-10-128-714-3127	Sequence 3127, Ap
12	33	86.8	555	14	US-10-128-714-8127	Sequence 8127, Ap
13	33	86.8	724	15	US-10-369-493-21897	Sequence 21897, A
14	31	81.6	273	12	US-10-424-599-163220	Sequence 163220,
15	31	81.6	462	12	US-10-425-114-42513	Sequence 42513, A
16	31	81.6	4675	15	US-10-093-463-74	Sequence 74, Appl
17	31	81.6	4691	15	US-10-093-463-72	Sequence 72, Appl
18	30	78.9	8	10	US-09-909-062-125	Sequence 125, App
19	30	78.9	8	10	US-09-909-062-127	Sequence 127, App
20	30	78.9	14	9	US-09-747-419-31	Sequence 31, Appl
21	30	78.9	14	14	US-10-259-275-31	Sequence 31, Appl
22	30	78.9	18	14	US-10-300-757-10	Sequence 10, Appl
23	30	78.9	20	10	US-09-775-052-50	Sequence 50, Appl
24	30	78.9	20	12	US-10-232-884-19	Sequence 19, Appl
25	30	78.9	94	14	US-10-300-757-22	Sequence 22, Appl
26	30	78.9	104	9	US-09-864-761-39527	Sequence 39527, A
27	30	78.9	248	15	US-10-369-493-14264	Sequence 14264, A
28	30	78.9	248	15	US-10-369-493-14770	Sequence 14770, A
29	30	78.9	257	15	US-10-369-493-11482	Sequence 11482, A
30	30	78.9	259	12	US-10-424-599-156429	Sequence 156429,
31	30	78.9	331	9	US-09-815-242-11909	Sequence 11909, A
32	30	78.9	377	12	US-10-424-599-222816	Sequence 222816,
33	30	78.9	379	12	US-10-424-599-278432	Sequence 278432,
34	30	78.9	427	14	US-10-156-761-9822	Sequence 9822, Ap
35	30	78.9	470	12	US-10-282-122A-49279	Sequence 49279, A
36	30	78.9	505	15	US-10-365-620-62	Sequence 62, Appl
37	30	78.9	677	12	US-10-425-114-55538	Sequence 55538, A
38	30	78.9	729	15	US-10-365-620-64	Sequence 64, Appl
39	30	78.9	764	14	US-10-300-757-20	Sequence 20, Appl
40	30	78.9	774	12	US-10-282-122A-62240	Sequence 62240, A
41	30	78.9	809	15	US-10-369-493-1573	Sequence 1573, Ap
42	30	78.9	896	14	US-10-210-296-5	Sequence 5, Appli
43	30	78.9	1040	9	US-09-929-955-9	Sequence 9, Appli
44	30	78.9	1040	13	US-10-104-966-9	Sequence 9, Appli
45	30	78.9	1070	12	US-10-425-114-57815	Sequence 57815, A

ALIGNMENTS

RESULT 1
US-09-777-785-1
; Sequence 1, Application US/09777785
; Patent No. US20020103135A1
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; TITLE OF INVENTION: Azapeptides Useful In The Treatment Of Hepatitis C
; FILE REFERENCE: IN01130K1 US
; CURRENT APPLICATION NUMBER: US/09/777,785
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/181,017
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:azapeptide
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: UNSURE
; LOCATION: (7)
; OTHER INFORMATION: 2[(4-nitrophenoxy)carbonyl]-2-propylhydrazine
US-09-777-785-1

Query Match 97.4%; Score 37; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7
```

RESULT 2

```
US-09-777-785-3
; Sequence 3, Application US/09777785
; Patent No. US20020103135A1
; GENERAL INFORMATION:
;  APPLICANT: Zhang, Rumin
;  TITLE OF INVENTION: Azapeptides Useful In The Treatment Of Hepatitis C
;  FILE REFERENCE: IN01130K1 US
;  CURRENT APPLICATION NUMBER: US/09/777,785
;  CURRENT FILING DATE:  2001-02-06
;  PRIOR APPLICATION NUMBER: 60/181,017
;  PRIOR FILING DATE: 2000-02-08
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:azapeptide
;   NAME/KEY: MOD_RES
;   LOCATION: (1)
;   OTHER INFORMATION: ACETYLATION
;   NAME/KEY: UNSURE
;   LOCATION: (7)
;   OTHER INFORMATION: 2-[(1,2,2,2-tetrachloroethoxy)carbonyl]-2-propylhy
;   OTHER INFORMATION: drazine
US-09-777-785-3

  Query Match             97.4%;  Score 37;  DB 9;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 3
US-09-909-062-128
; Sequence 128, Application US/09909062
; Publication No. US20030036501A1
; GENERAL INFORMATION:
;  APPLICANT: Saksena, Anil K
;  APPLICANT:  Girijavaliabhan, Viyyor M
;  APPLICANT:  Lovey, Raymond G
;  APPLICANT:  Jao, Edwin
;  APPLICANT:  Bennett, Frank
;  APPLICANT:  McCormick, Jinping L
;  APPLICANT:  Pike, Russell E
;  APPLICANT:  Bogen, Stephane L
;  APPLICANT:  Liu, Yi-Tsung
;  APPLICANT:  Arasappan, Ashok
;  APPLICANT:  Pinto, Patrick A
;  APPLICANT:  Njoroge, F George
;  APPLICANT:  Ganguly, Ashit
;  TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C V
;  FILE REFERENCE: IN01157K-US
;  CURRENT APPLICATION NUMBER: US/09/909,062
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: 60/220,109
;  PRIOR FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 149
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 128
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: ACETYLATION
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (8)..(8)
;   OTHER INFORMATION: norvaline
US-09-909-062-128

  Query Match             97.4%;  Score 37;  DB 10;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||||
Db          1 DTEDVVP 7


RESULT 4
US-10-282-122A-45763
; Sequence 45763, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45763
;   LENGTH: 2358
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
US-10-282-122A-45763

  Query Match             89.5%;  Score 34;  DB 12;  Length 2358;
  Best Local Similarity   85.7%;  Pred. No. 6.9e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||:|
Db         92 DTEDVLP 98
```

```
RESULT 5
US-10-424-599-197370
; Sequence 197370, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 197370
;   LENGTH: 72
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(72)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_20251C.1.pep
US-10-424-599-197370

  Query Match             86.8%;  Score 33;  DB 12;  Length 72;
  Best Local Similarity   85.7%;  Pred. No. 22;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||||| |
Db         55 DTEDVTP 61


RESULT 6
US-10-425-114-47171
; Sequence 47171, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 47171
;   LENGTH: 372
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3732-056-A5_FLI.pep
US-10-425-114-47171

  Query Match             86.8%;  Score 33;  DB 12;  Length 372;
  Best Local Similarity   85.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:|||||
Db        305 DSEDVVP 311


RESULT 7
US-10-425-114-57445
; Sequence 57445, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57445
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3732-012-D10_FLI.pep
US-10-425-114-57445

  Query Match             86.8%;  Score 33;  DB 12;  Length 374;
  Best Local Similarity   85.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:|||||
Db        307 DSEDVVP 313


RESULT 8
US-10-425-114-72029
; Sequence 72029, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Asscciated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72029
;   LENGTH: 386
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3245-328-E3_FLI.pep
US-10-425-114-72029

  Query Match             86.8%;  Score 33;  DB 12;  Length 386;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:|||||
Db        319 DSEDVVP 325


RESULT 9
US-10-425-114-72218
; Sequence 72218, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
```

```
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72218
;   LENGTH: 386
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700048277_FLI.pep
US-10-425-114-72218

  Query Match             86.8%;  Score 33;  DB 12;  Length 386;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:|||||
Db        319 DSEDVVP 325


RESULT 10
US-10-238-075-903
; Sequence 903, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
;  APPLICANT: I.N.S.E.R.M.
;  TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
;  TITLE OF INVENTION:  E.coli, and biological uses of these polynucleotides and of thei
;  FILE REFERENCE: BLANDINE
;  CURRENT APPLICATION NUMBER: US/10/238,075
;  CURRENT FILING DATE:  2002-09-10
;  PRIOR APPLICATION NUMBER: 0003145
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 1576
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 903
;   LENGTH: 415
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-10-238-075-903

  Query Match             86.8%;  Score 33;  DB 14;  Length 415;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||||::|
Db         79 DTEDILP 85


RESULT 11
US-10-128-714-3127
; Sequence 3127, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3127
;   LENGTH: 548
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-3127

  Query Match             86.8%;  Score 33;  DB 14;  Length 548;
  Best Local Similarity   85.7%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:|||||
Db         75 DSEDVVP 81


RESULT 12
US-10-128-714-8127
; Sequence 8127, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8127
;   LENGTH: 555
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-8127

  Query Match             86.8%;  Score 33;  DB 14;  Length 555;
  Best Local Similarity   85.7%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:|||||
Db         75 DSEDVVP 81


RESULT 13
US-10-369-493-21897
; Sequence 21897, Application US/10369493
```

; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21897
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-21897

Query Match 86.8%; Score 33; DB 15; Length 724;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
||:||||
Db 630 DTKDVVP 636

RESULT 14
US-10-424-599-163220
; Sequence 163220, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 163220
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_118406C.1.pep
US-10-424-599-163220

Query Match 81.6%; Score 31; DB 12; Length 273;
Best Local Similarity 85.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
| |||||
Db 61 DIEDVVP 67

RESULT 15
US-10-425-114-42513
; Sequence 42513, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 42513
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700263159_FLI.pep
US-10-425-114-42513

Query Match 81.6%; Score 31; DB 12; Length 462;
Best Local Similarity 85.7%; Pred. No. 4.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
| |||||
Db 135 DEEDVVP 141

Search completed: March 31, 2004, 16:52:58
Job time : 35.4 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:41:17 ; Search time 11.2 Seconds
(without alignments)
68.708 Million cell updates/sec

Title: US-09-909-077-4
Perfect score: 38
Sequence: 1 DTEDVVPX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	86.8	724	1	S53934	glutamate-tRNA lig
2	32	84.2	437	2	T08094	probable sulfate a
3	32	84.2	580	2	B84554	hypothetical prote
4	32	84.2	1323	2	AH0225	1-pyrroline-5-carb
5	31	81.6	324	2	S58142	coat protein - pha
6	31	81.6	393	2	B84758	probable katanin [
7	31	81.6	637	2	T04552	hypothetical prote
8	31	81.6	762	2	JC7174	N,N-dimethylformam
9	31	81.6	1136	2	A56559	enhancer-trap-locu
10	31	81.6	3898	1	GNWVHC	genome polyprotein
11	30	78.9	148	1	QQBE25	BERF3 protein - hu
12	30	78.9	158	2	G70323	conserved hypothet
13	30	78.9	273	2	AB3093	dehydrogenase/redu
14	30	78.9	273	2	H98193	probable short-cha
15	30	78.9	331	2	H83240	probable ATP-bindi
16	30	78.9	392	2	T32555	hypothetical prote
17	30	78.9	412	2	T15214	hypothetical prote
18	30	78.9	424	1	A36000	sperm-binding glyc
19	30	78.9	439	2	G88103	protein W10G11.17
20	30	78.9	495	2	T00811	hypothetical prote
21	30	78.9	544	2	T26559	hypothetical prote
22	30	78.9	568	2	S42225	major envelope gly
23	30	78.9	585	2	G86200	protein F12K11.15
24	30	78.9	653	2	AI3404	transposase BMEI12
25	30	78.9	809	2	S67665	ubiquitin-specific
26	30	78.9	992	2	A31666	hypothetical prote
27	30	78.9	1171	2	S57829	genome polyprotein
28	30	78.9	1311	2	T33757	hypothetical prote
29	30	78.9	2135	2	T14602	variant-specific s
30	30	78.9	2517	2	S58380	probable RNA-direc
31	30	78.9	3011	1	GNWVCH	genome polyprotein
32	30	78.9	3898	1	GNWVHB	genome polyprotein
33	30	78.9	3898	2	S57437	genome polyprotein
34	30	78.9	3898	2	S58295	polyprotein - hog
35	29	76.3	71	2	T18037	hypothetical prote
36	29	76.3	172	2	F97320	acetyltransferase
37	29	76.3	204	2	D86048	hypothetical prote
38	29	76.3	213	2	S35522	heterochromatin pr
39	29	76.3	231	2	H91201	hypothetical prote
40	29	76.3	295	2	AE1398	transcription regu
41	29	76.3	295	2	AH1773	transcription regu
42	29	76.3	343	2	H91261	probable membrane
43	29	76.3	343	2	D86102	probable membrane
44	29	76.3	343	2	A65217	hypothetical 36.9
45	29	76.3	344	2	T34421	hypothetical prote

ALIGNMENTS

RESULT 1
S53934
glutamate-tRNA ligase (EC 6.1.1.17) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G0583; protein HRB724; protein NRC145; protein YGL245w
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
C;Accession: S53934; S59351; S60484; S61616; S64270; S64271
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, April 1995
A;Description: The sequence of a 11.1 kb DNA fragment between ADH4 and ADE5 on the left
A;Reference number: S53934
A;Accession: S53934
A;Molecule type: DNA
A;Residues: 1-724 <VAN>
A;Cross-references: EMBL:Z49149; NID:g793865; PIDN:CAA89009.1; PID:g793866
R;Frantz, J.D.; Gilbert, W.
submitted to the EMBL Data Library, July 1995
A;Description: Isolation and sequence characterization of the gene encoding the yeast c
A;Reference number: S59351
A;Accession: S59351
A;Molecule type: DNA
A;Residues: 1-224,'D',226-488,'A',490-525,'S',527-561,'M',563-713,'VNLSTSMVQRNKHHISNVYT
A;Cross-references: EMBL:U32265; NID:g1008482; PIDN:AAA78905.1; PID:g1008483
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
Yeast 11, 1519-1523, 1995
A;Title: The sequence of an 11.1 kb DNA fragment between ADH4 and ADE5 on the left arm
A;Reference number: S60484; MUID:96353434; PMID:8750240
A;Accession: S60484
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 57-724 <VAW>
A;Cross-references: EMBL:Z49149
R;Coissac, E.; Maillier, E.; Robineau, S.; Netter, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: S61598
A;Accession: S61616
A;Molecule type: DNA
A;Residues: 1-145 <COI>
A;Cross-references: EMBL:X94357; NID:g1150575; PIDN:CAA64142.1; PID:g1150594
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64263
A;Accession: S64270
A;Molecule type: DNA
A;Residues: 1-724 <VAF>
A;Cross-references: EMBL:Z72767; NID:g1945311; PIDN:CAA96964.1; PID:g1322915; GSPDB:GN0
A;Experimental source: strain S288C
R;Coissac, E.; Maillier, E.; Netter, P.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64271
A;Accession: S64271
A;Molecule type: DNA

A;Residues: 1-145 <COW>
A;Cross-references: EMBL:Z72767; GSPDB:GN00007; MIPS:YGL245w
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YGL245w
A;Cross-references: SGD:S0003214
A;Map position: 7L
C;Superfamily: yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis
F;218-495/Domain: glutamine-tRNA ligase homology <EGL>

Query Match 86.8%; Score 33; DB 1; Length 724;
Best Local Similarity 85.7%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
||:||||
Db 630 DTKDVVP 636

RESULT 2
T08094
probable sulfate adenylyltransferase (EC 2.7.7.4) ATS1 - Chlamydomonas reinhardtii
N;Alternate names: ATP sulfurylase
C;Species: Chlamydomonas reinhardtii
C;Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 03-Dec-1999
C;Accession: T08094
R;Yildiz, F.H.; Davies, J.P.; Grossman, A.
submitted to the EMBL Data Library, April 1996
A;Description: Controlled expression of the ATP sulfurylase gene of Chlamydomonas reinha
A;Reference number: Z16349
A;Accession: T08094
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-437 <YIL>
A;Cross-references: EMBL:U57088; NID:g1336212; PIDN:AAB01234.1; PID:g1336213
A;Experimental source: strain CC125
C;Genetics:
A;Gene: ATS1
C;Function:
A;Description: catalyzes the reaction of sulfate and ATP to form adenylylsulfate and pyr
C;Superfamily: sulfate adenylyltransferase met3-1; sulfate adenylyltransferase homology
C;Keywords: nucleotidyltransferase
F;56-434/Domain: sulfate adenylyltransferase homology <SAT>

Query Match 84.2%; Score 32; DB 2; Length 437;
Best Local Similarity 71.4%; Pred. No. 47;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
|:||:||
Db 138 DSEDIVP 144

RESULT 3
B84554
hypothetical protein At2g17600 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84554
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-580 <STO>
A;Cross-references: GB:AE002093; NID:g4914381; PIDN:AAD32917.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g17600
A;Map position: 2

Query Match 84.2%; Score 32; DB 2; Length 580;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
:|||:||
Db 266 ETEDIVP 272

RESULT 4
AH0225
1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12) [imported] - Yersinia pestis (stra
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C;Accession: AH0225
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0225
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1323 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90668.1; PID:g15979873; GSPDB:GN00175
C;Genetics:
A;Gene: putA
C;Superfamily: bifunctional protein putA
C;Keywords: oxidoreductase

Query Match 84.2%; Score 32; DB 2; Length 1323;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7
|||| :|
Db 63 DTEDAIP 69

RESULT 5
S58142
coat protein - phage SPP1
C;Species: phage SPP1
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 11-May-2000
C;Accession: S58142; T42283
R;Becker, B.; Gassel, M.; Tavares, P.; Lurz, R.; Alonso, J.C.
submitted to the EMBL Data Library, July 1995
A;Description: Head morphogenesis of the Bacillus subtilis bacteriophage SPP1.
A;Reference number: S58137
A;Accession: S58142
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-324 <BEC>
A;Cross-references: EMBL:X89721; NID:g1052805; PIDN:CAA61870.1; PID:g1052811
R;Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.
Gene 204, 201-212, 1997
A;Title: The complete nucleotide sequence and functional organization of Bacillus subtil
A;Reference number: Z22137; MUID:98094274; PMID:9434185
A;Accession: T42283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-324 <ALO>
A;Cross-references: EMBL:X97918; PIDN:CAA66544.1
C;Superfamily: major capsid protein gpG

Query Match 81.6%; Score 31; DB 2; Length 324;
Best Local Similarity 71.4%; Pred. No. 55;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVP 7

```
              ||:|:||
Db          71 DTDDLVP 77
```

RESULT 6
B84758
probable katanin [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84758
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84758
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-393 <STO>
A;Cross-references: GB:AE002093; NID:g3128218; PIDN:AAC26698.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g34560
A;Map position: 2

```
  Query Match             81.6%;  Score 31;  DB 2;  Length 393;
  Best Local Similarity   85.7%;  Pred. No. 69;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | |||||
Db        341 DREDVVP 347
```

RESULT 7
T04552
hypothetical protein F28J12.220 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04552
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377
A;Accession: T04552
A;Molecule type: DNA
A;Residues: 1-637 <BEV>
A;Cross-references: EMBL:AL021710
A;Experimental source: cultivar Columbia; BAC clone F28J12
C;Genetics:
A;Map position: 4
A;Introns: 110/3; 394/3; 438/3; 511/2; 548/3; 594/3
A;Note: F28J12.220

```
  Query Match             81.6%;  Score 31;  DB 2;  Length 637;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | |||||
Db        420 DIEDVVP 426
```

RESULT 8
JC7174
N,N-dimethylformamidase (EC 3.5.1.56) beta chain - Alcaligenes sp.
C;Species: Alcaligenes sp.
C;Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
C;Accession: JC7174; PC7051
R;Hasegawa, Y.; Tokuyama, T.; Iwaki, H.
Biosci. Biotechnol. Biochem. 63, 2091-2096, 1999
A;Title: Cloning and expression of the N,N-dimethylformamidase gene from Alcaligenes sp.
A;Reference number: JC7173; MUID:20128227; PMID:10664842
A;Accession: JC7174
A;Molecule type: DNA
A;Residues: 1-762 <HAS>
A;Cross-references: DDBJ:AB028874
A;Experimental source: strain KUFA-1
A;Accession: PC7051
A;Molecule type: protein
A;Residues: 1-25 <HA2>
C;Comment: This enzyme is an amidase used in the chemical industry as a solvent for pol
C;Genetics:
A;Gene: dmfA2
C;Keywords: hydrolase

```
  Query Match             81.6%;  Score 31;  DB 2;  Length 762;
  Best Local Similarity   71.4%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||| :|
Db        367 DTEDYIP 373
```

RESULT 9
A56559
enhancer-trap-locus-1 protein - mouse (fragment)
N;Alternate names: homeotic regulator brahma protein homolog Etl-1
C;Species: Mus musculus (house mouse)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
C;Accession: A56559; S31583
R;Soininen, R.; Schoor, M.; Henseling, U.; Tepe, C.; Kisters-Woike, B.; Rossant, J.; Go
Mech. Dev. 39, 111-123, 1992
A;Title: The mouse Enhancer trap locus 1 (Etl-1): a novel mammalian gene related to Dro
A;Reference number: A56559; MUID:93144171; PMID:1489724
A;Accession: A56559
A;Molecule type: mRNA
A;Residues: 1-1136 <SOI>
A;Cross-references: EMBL:X69942; NID:g50865; PIDN:CAA49560.1; PID:g50866
A;Note: sequence extracted from NCBI backbone (NCBIP:123884)

```
  Query Match             81.6%;  Score 31;  DB 2;  Length 1136;
  Best Local Similarity   85.7%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||||| |
Db        213 DTEDVSP 219
```

RESULT 10
GNWVHC
genome polyprotein - hog cholera virus (strain Alfort)
C;Species: hog cholera virus
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
C;Accession: A34037
R;Meyers, G.; Ruemenapf, T.; Thiel, H.J.
Virology 171, 555-567, 1989
A;Title: Molecular cloning and nucleotide sequence of the genome of hog cholera virus.
A;Reference number: A34037; MUID:89348014; PMID:2763466
A;Accession: A34037
A;Molecule type: genomic RNA
A;Residues: 1-3898 <MEY>
A;Cross-references: GB:J04358; NID:g325462; PIDN:AAA43844.1; PID:g325463
C;Superfamily: pestivirus genome polyprotein
C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop; polyprotein; transmembrane p
F;2-231/Product: viral proteinase p20 #status predicted <VPT>
F;545-1111/Product: major envelope glycoprotein gp55 #status predicted <EGP>
F;1815-1822/Region: nucleotide-binding motif A (P-loop)
F;1906-1911/Region: nucleotide-binding motif B
F;1910-1913/Region: DEXH motif
F;157,269,274,278,293,362,367,410,425,500,594,805,810,918,949,986,1713,2134,2217,2494,2

```
  Query Match             81.6%;  Score 31;  DB 1;  Length 3898;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 TEDVVP 7
              ||||||
Db        601 TEDVVP 606


RESULT 11
QQBE25
BERF3 protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C;Accession: E43042; A03766; S33014
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A;Reference number: A93065; MUID:85035713; PMID:6092825
A;Accession: E43042
A;Molecule type: DNA
A;Residues: 1-148 <BAN>
A;Cross-references: EMBL:V01555
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667; PMID:6087149
A;Contents: annotation; protein coding region
C;Superfamily: human herpesvirus 4 nuclear antigen EBNA-3C

  Query Match             78.9%;  Score 30;  DB 1;  Length 148;
  Best Local Similarity   85.7%;  Pred. No. 37;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||| ||
Db        105 DTEDNVP 111


RESULT 12
G70323
conserved hypothetical protein aq_260 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C;Accession: G70323
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: G70323
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-158 <AQF>
A;Cross-references: GB:AE000681; NID:g2982963; PIDN:AAC06588.1; PID:g2982974; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_260
C;Superfamily: nus operon 15K protein

  Query Match             78.9%;  Score 30;  DB 2;  Length 158;
  Best Local Similarity   57.1%;  Pred. No. 40;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | ||::|
Db         68 DVEDIIP 74


RESULT 13
AB3093
dehydrogenase/reductase sdhX [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AB3093
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB3093
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45160.1; PID:g17742835; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: sdhX
A;Map position: linear chromosome
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

  Query Match             78.9%;  Score 30;  DB 2;  Length 273;
  Best Local Similarity   71.4%;  Pred. No. 75;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | ||:||
Db        241 DIEDIVP 247


RESULT 14
H98193
probable short-chain dehydrogenase PA2918 [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98193
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98193
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89074.1; PID:g15158872; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1000
A;Map position: linear chromosome
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

  Query Match             78.9%;  Score 30;  DB 2;  Length 273;
  Best Local Similarity   71.4%;  Pred. No. 75;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | ||:||
Db        241 DIEDIVP 247


RESULT 15
H83240
probable ATP-binding component of ABC transporter PA3254 [imported] - Pseudomonas aerugi
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83240
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83240
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-331 <STO>
```

A;Cross-references: GB:AE004747; GB:AE004091; NID:g9949362; PIDN:AAG06642.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3254

Query Match 78.9%; Score 30; DB 2; Length 331;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 DTEDVVP 7
              | |||||
Db         63 DGEDVVP 69
```

Search completed: March 31, 2004, 16:49:33
Job time : 12.2 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:40 ; Search time 7.46667 Seconds
(without alignments)
55.789 Million cell updates/sec

Title: US-09-909-077-4
Perfect score: 38
Sequence: 1 DTEDVVPX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	86.8	724	1	SYEC_YEAST	P46655 saccharomyc
2	31	81.6	1021	1	SRD1_MOUSE	Q04692 mus musculu
3	31	81.6	3898	1	POLG_HCVA	P19712 hog cholera
4	30	78.9	158	1	Y260_AQUAE	O66619 aquifex aeo
5	30	78.9	237	1	THYX_BPPHC	Q9zx92 bacteriopha
6	30	78.9	424	1	ZP3_CALSQ	P53786 callithrix
7	30	78.9	424	1	ZP3_HUMAN	P21754 homo sapien
8	30	78.9	424	1	ZP3_MACRA	P53785 macaca radi
9	30	78.9	455	1	DALD_KLEPN	O52720 klebsiella
10	30	78.9	786	1	MUS2_CLOTE	Q891u1 clostridium
11	30	78.9	809	1	UBP1_YEAST	P25037 saccharomyc
12	30	78.9	992	1	EBN6_EBV	P03204 epstein-bar
13	30	78.9	3011	1	POLG_HCVH	P27958 h genome po
14	30	78.9	3898	1	POLG_HCVB	P21530 hog cholera
15	29	76.3	213	1	HP1_DROVI	P29227 drosophila
16	29	76.3	343	1	YJCR_ECOLI	P32716 escherichia
17	29	76.3	352	1	DIN1_SCHPO	O13836 schizosacch
18	29	76.3	371	1	VA06_VACCV	P29192 vaccinia vi
19	29	76.3	372	1	VA06_VACCC	P20985 vaccinia vi
20	29	76.3	372	1	VA06_VARV	P33833 variola vir
21	29	76.3	459	1	TRME_STAEP	Q8cmn5 staphylococ
22	29	76.3	462	1	ACDG_METKA	Q8txf1 methanopyru
23	29	76.3	484	1	FLHF_CAMJE	O52908 campylobact
24	29	76.3	654	1	YPJ1_CAEEL	P48322 caenorhabdi
25	29	76.3	713	1	HS90_EIMTE	O44001 eimeria ten
26	29	76.3	834	1	GYRA_CHLPN	Q9z8r4 chlamydia p
27	29	76.3	854	1	KDPD_RATRA	O34971 rathayibact
28	29	76.3	993	1	V1A_CMVFN	P17769 cucumber mo
29	29	76.3	993	1	V1A_CMVII	Q83270 cucumber mo
30	29	76.3	993	1	V1A_CMVIX	Q66121 cucumber mo
31	29	76.3	993	1	V1A_CMVO	P20122 cucumber mo
32	29	76.3	993	1	V1A_CMVY	Q83264 cucumber mo
33	29	76.3	1442	1	YJ9F_YEAST	P47169 saccharomyc
34	29	76.3	2672	1	GCN1_YEAST	P33892 saccharomyc
35	29	76.3	4499	1	DYHA_CHLRE	Q39610 chlamydomon
36	28	73.7	156	1	L18B_LUPLU	P52779 lupinus lut
37	28	73.7	266	1	ITMB_HUMAN	Q9y287 homo sapien
38	28	73.7	310	1	E13A_HORVU	P34742 hordeum vul
39	28	73.7	320	1	FLA1_TREHY	P32520 treponema h
40	28	73.7	359	1	TMO1_HUMAN	P28289 homo sapien
41	28	73.7	359	1	TMO1_MOUSE	P49813 mus musculu
42	28	73.7	359	1	TMO1_RAT	P70567 rattus norv
43	28	73.7	366	1	HNLS_SORBI	P52708 sorghum bic
44	28	73.7	384	1	DPO4_THETN	P58965 thermoanaer
45	28	73.7	474	1	VTDB_HUMAN	P02774 homo sapien

ALIGNMENTS

RESULT 1
SYEC_YEAST
ID SYEC_YEAST STANDARD; PRT; 724 AA.
AC P46655;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Glutamyl-tRNA synthetase, cytoplasmic (EC 6.1.1.17) (Glutamate--tRNA
DE ligase) (GluRS) (P85).
GN YGL245W OR G0583 OR HRB724.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Frantz J.D., Gilbert W.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 1-146 FROM N.A.
RC STRAIN=S288c / FY1679;
RA Coissac E., Maillier E., Netter P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC ----------

CC ----------
DR EMBL; U32265; AAA78905.1; -.
DR EMBL; Z49149; CAA89009.1; -.
DR EMBL; Z72767; CAA96964.1; -.
DR EMBL; X94357; CAA64142.1; -.
DR PIR; S53934; S53934.
DR HSSP; P00962; 1GTR.
DR GermOnline; 141294; -.
DR SGD; S0003214; YGL245W.
DR InterPro; IPR004526; GltX_arch.
DR InterPro; IPR000924; Glu_tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_I.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c_C; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRFAMs; TIGR00463; gltX_arch; 1.

```
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   SITE        226    235       "HIGH" REGION.
FT   SITE        453    457       "KMSKS" REGION.
FT   BINDING     456    456       ATP (BY SIMILARITY).
FT   CONFLICT    225    225       E -> D (IN REF. 1).
FT   CONFLICT    489    489       V -> A (IN REF. 1).
FT   CONFLICT    526    526       P -> S (IN REF. 1).
FT   CONFLICT    562    562       V -> M (IN REF. 1).
FT   CONFLICT    714    724       GKSVNKYGAKK -> VNLSTSMVQRNKHHISNVYTYLCYFS
FT                                TSTF (IN REF. 1).
SQ   SEQUENCE   724 AA;  82662 MW;  34669BFB69CD41BE CRC64;

  Query Match             86.8%;  Score 33;  DB 1;  Length 724;
  Best Local Similarity   85.7%;  Pred. No. 22;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||:||||
Db        630 DTKDVVP 636


RESULT 2
SRD1_MOUSE
ID   SRD1_MOUSE     STANDARD;      PRT;  1021 AA.
AC   Q04692;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   SWI/SNF-related, matrix associated, actin-dependent regulator of
DE   chromatin subfamily A containing DEAD/H box 1 (Enhancer trap locus
DE   homolog 1) (Etl-1).
GN   SMARCAD1 OR ETL1 OR KIAA1122.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=C57BL/6; TISSUE=Embryo;
RX   MEDLINE=93144171; PubMed=1489724;
RA   Soininen R., Schoor M., Henseling U., Tepe C., Kisters-Woike B.,
RA   Rossant J., Gossler A.;
RT   "The mouse enhancer trap locus 1 (ETL-1) gene: a novel mammalian gene
RT   related to transcriptional regulators in Drosophila and yeast.";
RL   Mech. Dev. 39:111-123(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=FVB/N;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SUBCELLULAR LOCATION, AND DEVELOPMENTAL STAGE.
RX   MEDLINE=94033710; PubMed=8219362;
RA   Schoor M., Schuster-Gossler K., Gossler A.;
RT   "The Etl-1 gene encodes a nuclear protein differentially expressed
RT   during early mouse development.";
RL   Dev. Dyn. 197:227-237(1993).
CC   -!- FUNCTION: Probable ATP-dependent DNA helicase.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q04692-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q04692-2; Sequence=VSP_007080;
CC         Note=No experimental confirmation available;
CC   -!- DEVELOPMENTAL STAGE: Detected at low levels in fertilized and
CC       unfertilized eggs. Levels increased in two-cell embryos, decreased
CC       up to morula stage and were highest in blastocysts. Highly
CC       expressed in the inner cell mass of 3.5 day old blastocysts.
CC       Highly expressed in exctoderm and visceral endoderm at day 5.5.
CC       Detected througout the brain and spinal cord at day 10 to 15.
CC       Detected in the basal layer of the epidermis after day 12.5, in
CC       particular on snout and distal on fore- and hind-limbs.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69942; CAA49560.1; ALT_INIT.
DR   EMBL; AK122454; BAC65736.1; ALT_INIT.
DR   EMBL; BC042442; AAH42442.1; -.
DR   PIR; A56559; A56559.
DR   MGD; MGI:95453; Smarcad1.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hydrolase; Helicase; ATP-binding; Nuclear protein;
KW   Alternative splicing.
FT   NP_BIND     516    524       ATP (BY SIMILARITY).
FT   NP_BIND     892    899       ATP (BY SIMILARITY).
FT   SITE        623    626       DEGH BOX.
FT   SITE       1000   1003       DEAD BOX.
FT   DOMAIN      716    733       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   VARSPLIC      1    185       Missing (in isoform 2).
FT                                /FTId=VSP_007080.
FT   CONFLICT    857    857       I -> S (IN REF. 1).
SQ   SEQUENCE   1021 AA;  116450 MW;  E3237AA2B135538A CRC64;

  Query Match             81.6%;  Score 31;  DB 1;  Length 1021;
  Best Local Similarity   85.7%;  Pred. No. 90;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||||| |
Db         98 DTEDVSP 104
```

```
RESULT 3
POLG_HCVA
ID   POLG_HCVA      STANDARD;      PRT;  3898 AA.
AC   P19712;
DT   01-FEB-1991 (Rel. 17, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polyprotein.
OS   Hog cholera virus (strain Alfort) (Swine fever virus).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Pestivirus.
OX   NCBI_TaxID=11097;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89348014; PubMed=2763466;
RA   Meyers G., Ruemenapf T., Thiel H.-J.;
RT   "Molecular cloning and nucleotide sequence of the genome of hog
RT   cholera virus.";
RL   Virology 171:555-567(1989).
RN   [2]
RP   REVISION TO 2731.
RA   Meyers G.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN
CC       WITH HELICASE AND PROTEASE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: THE GP51-GP54 PROTEIN IS ANCHORED TO THE
CC       VIRAL ENVELOPE.
CC   -!- SIMILARITY: TO BOVINE VIRAL DIARRHEA VIRUS GENOME POLYPROTEIN.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04358; AAA43844.2; -.
DR   HSSP; P27958; 1A1V.
DR   MEROPS; C53.001; -.
DR   MEROPS; S31.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR008751; Peptidase_C53.
DR   InterPro; IPR000280; Peptidase_S31.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   InterPro; IPR001568; RNase_T2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF05550; Peptidase_C53; 1.
DR   Pfam; PF05578; Peptidase_S31; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   PRINTS; PR00729; CDVENDOPTASE.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00531; RNASE_T2_2; UNKNOWN_1.
KW   Polyprotein; Glycoprotein; Transmembrane; Hydrolase; Serine protease;
KW   Helicase.
FT   CHAIN        ?1   ?267       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN      ?268   ?500       GP44-GP48 (E2) (POTENTIAL).
FT   CHAIN      ?501   ?689       GP33 (E3) (POTENTIAL).
FT   CHAIN      ?690  ?1060       GP55 (ENVELOPE PROTEIN E1).
FT   CHAIN     ?1611  ?2111       P80 (POTENTIAL).
FT   TRANSMEM   1032   1048       POTENTIAL.
FT   ACT_SITE   1658   1658       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1695   1695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1752   1752       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    332    332       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    500    500       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    805    805       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    810    810       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    949    949       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    986    986       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1713   1713       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2134   2134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2217   2217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2494   2494       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2787   2787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2815   2815       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2891   2891       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3211   3211       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3316   3316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3689   3689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3698   3698       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3794   3794       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     387    387       T -> A.
FT   VARIANT    3542   3542       R -> S.
SQ   SEQUENCE   3898 AA;  438570 MW;  2C1F17B8A359D0F6 CRC64;

  Query Match             81.6%;  Score 31;  DB 1;  Length 3898;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||||
Db        601 TEDVVP 606


RESULT 4
Y260_AQUAE
ID   Y260_AQUAE     STANDARD;      PRT;   158 AA.
AC   O66619;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical UPF0090 protein AQ_260.
GN   AQ_260.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- SIMILARITY: Belongs to the UPF0090 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AE000681; AAC06588.1; -.
DR   PIR; G70323; G70323.
DR   HAMAP; MF_01077; -; 1.
DR   InterPro; IPR003728; DUF150.
DR   Pfam; PF02576; DUF150; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   158 AA;  18147 MW;  6F9869AB382FE734 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 158;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | ||::|
Db         68 DVEDIIP 74


RESULT 5
THYX_BPPHC
ID   THYX_BPPHC     STANDARD;      PRT;   237 AA.
AC   Q9ZX92;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable thymidylate synthase (EC 2.1.1.148) (TS) (TSase) (GP16).
GN   16.
OS   Bacteriophage phi-C31.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   Lambda-like viruses.
OX   NCBI_TaxID=10719;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Norwich;
RX   MEDLINE=99162580; PubMed=10051617;
RA   Hendrix R.W., Smith M.C.M., Burns N., Ford M.E., Hatfull G.F.;
RT   "Evolutionary relationships among diverse bacteriophages and
RT   prophages: all the world's a phage.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:2192-2197(1999).
CC   -!- FUNCTION: Catalyzes the formation of dTMP and tetrahydrofolate
CC       from dUMP and methylenetetrahydrofolate (By similarity).
CC   -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + dUMP +
CC       FADH(2) = dTMP + tetrahydrofolate + FAD.
CC   -!- COFACTOR: Binds 1 FAD per subunit (By similarity).
CC   -!- SIMILARITY: Belongs to the thymidylate synthase thyX family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ006589; CAA07140.1; -.
DR   InterPro; IPR003669; Thy1.
DR   Pfam; PF02511; Thy1; 1.
KW   Transferase; Methyltransferase; Nucleotide biosynthesis; FAD;
KW   Flavoprotein.
SQ   SEQUENCE   237 AA;  26016 MW;  3FDEF9BA6DF302C0 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 237;
  Best Local Similarity   85.7%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||| |||
Db        118 DTEPVVP 124


RESULT 6
ZP3_CALSQ
ID   ZP3_CALSQ      STANDARD;      PRT;   424 AA.
AC   P53786;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor).
GN   ZP3.
OS   Callithrix sp. (Marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC   Callithrix.
OX   NCBI_TaxID=9485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=94363314; PubMed=8081814;
RA   Thillai-Koothan P., van Duin M., Aitken R.J.;
RT   "Cloning, sequencing and oocyte-specific expression of the marmoset
RT   sperm receptor protein, ZP3.";
RL   Zygote 1:93-101(1993).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S71825; AAB31866.1; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    388    408       POTENTIAL.
FT   DOMAIN      409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    307       ZP.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  46809 MW;  1DACBD03026C2739 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 7
ZP3_HUMAN
ID   ZP3_HUMAN      STANDARD;      PRT;   424 AA.
AC   P21754; Q06633;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
```

```
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor)
DE   (ZP3A/ZP3B).
GN   ZP3 OR ZP3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ZP3A).
RX   MEDLINE=90349545; PubMed=2385582;
RA   Chamberlin M.E., Dean J.;
RT   "Human homolog of the mouse sperm receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6014-6018(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS ZP3A AND ZP3B).
RC   TISSUE=Ovary;
RX   MEDLINE=93122771; PubMed=1478648;
RA   van Duin M., Polman J.E., Verkoelen C.C., Bunschoten H.,
RA   Meyerink J.H., Olijve W., Aitken R.J.;
RT   "Cloning and characterization of the human sperm receptor ligand ZP3:
RT   evidence for a second polymorphic allele with a different frequency
RT   in the Caucasian and Japanese populations.";
RL   Genomics 14:1064-1070(1992).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN WHICH
CC       ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=ZP3A;
CC         IsoId=P21754-1; Sequence=Displayed;
CC       Name=ZP3B;
CC         IsoId=P21754-2; Sequence=VSP_006949, VSP_006950;
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60504; AAA61336.1; -.
DR   EMBL; X56777; CAA40095.1; -.
DR   EMBL; A18567; CAA01398.1; -.
DR   PIR; A36000; A36000.
DR   Genew; HGNC:13189; ZP3.
DR   MIM; 182889; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family; Alternative splicing.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN      23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   388    408       POTENTIAL.
FT   DOMAIN     409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      45    307       ZP.
FT   CARBOHYD   125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   364    372       PLIFLDRRG -> ATDLPGQEW (in isoform ZP3B).
FT                               /FTId=VSP_006949.
FT   VARSPLIC   373    424       Missing (in isoform ZP3B).
FT                               /FTId=VSP_006950.
FT   CONFLICT   345    345       R -> T (IN REF. 2).
SQ   SEQUENCE   424 AA;  47028 MW;  94517B7B66EFFE06 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 8
ZP3_MACRA
ID   ZP3_MACRA      STANDARD;      PRT;   424 AA.
AC   P53785;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor).
GN   ZP3.
OS   Macaca radiata (Bonnet monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9548;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96249321; PubMed=8848588;
RA   Kolluri S.K., Kaul R., Banerjee K., Gupta S.K.;
RT   "Nucleotide sequence of cDNA encoding bonnet monkey (Macaca radiata)
RT   zona pellucida glycoprotein-ZP3.";
RL   Reprod. Fertil. Dev. 7:1209-1212(1995).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82639; CAA57961.1; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN      23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   388    408       POTENTIAL.
FT   DOMAIN     409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      45    307       ZP.
FT   CARBOHYD   125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  47040 MW;  3B41C4CFA3792331 CRC64;
```

```
  Query Match             78.9%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 9
DALD_KLEPN
ID   DALD_KLEPN     STANDARD;      PRT;   455 AA.
AC   O52720;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   D-arabinitol 4-dehydrogenase (EC 1.1.1.11).
GN   DALD.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1033-5P14 / KAY2026;
RX   MEDLINE=98304087; PubMed=9639934;
RA   Heuel H., Shakeri-Garakani A., Turgut S., Lengeler J.W.;
RT   "Genes for D-arabinitol and ribitol catabolism from Klebsiella
RT   pneumoniae.";
RL   Microbiology 144:1631-1639(1998).
CC   -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) = D-xylulose + NADH.
CC   -!- PATHWAY: D-arabinitol catabolism; first step.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: Belongs to the mannitol dehydrogenase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF045245; AAC26498.1; -.
DR   InterPro; IPR000669; Mannitol_dh.
DR   Pfam; PF01232; Mannitol_dh; 1.
DR   PRINTS; PR00084; MTLDHDRGNASE.
DR   PROSITE; PS00974; MANNITOL_DHGENASE; FALSE_NEG.
KW   Oxidoreductase; NAD.
SQ   SEQUENCE   455 AA;  51022 MW;  2D446E5232778DD5 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 455;
  Best Local Similarity   83.3%;  Pred. No. 60;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||:|
Db        311 TEDVIP 316


RESULT 10
MUS2_CLOTE
ID   MUS2_CLOTE     STANDARD;      PRT;   786 AA.
AC   Q891U1;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   MutS2 protein.
GN   MUTS2 OR MUTS OR CTC02274.
OS   Clostridium tetani.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC   -!- FUNCTION: Not known.
CC   -!- SIMILARITY: Belongs to the DNA mismatch repair mutS family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE015943; AAO36754.1; -.
DR   HAMAP; MF_00092; -; 1.
DR   InterPro; IPR000432; MutS_C.
DR   InterPro; IPR007696; MutS_III.
DR   InterPro; IPR002625; Smr/MutS2.
DR   Pfam; PF00488; MutS_V; 1.
DR   Pfam; PF01713; Smr; 1.
DR   PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; FALSE_NEG.
KW   ATP-binding; DNA-binding; Complete proteome.
FT   NP_BIND     332    339       ATP (POTENTIAL).
SQ   SEQUENCE   786 AA;  89223 MW;  7FB763E5C17F2683 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 786;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||| |||
Db        310 DTEKVVP 316


RESULT 11
UBP1_YEAST
ID   UBP1_YEAST     STANDARD;      PRT;   809 AA.
AC   P25037; Q07543;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Ubiquitin carboxyl-terminal hydrolase 1 (EC 3.1.2.15) (Ubiquitin
DE   thiolesterase 1) (Ubiquitin-specific processing protease 1)
DE   (Deubiquitinating enzyme 1).
GN   UBP1 OR YDL122W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91268082; PubMed=2050695;
RA   Tobias J.W., Varshavsky A.;
RT   "Cloning and functional analysis of the ubiquitin-specific protease
RT   gene UBP1 of Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 266:12021-12028(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Mueller-Auer S., Brueckner M., Schaefer M., Wagner G.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Has an ATP-independent isopeptidase activity, cleaving
CC       at the carboxyl terminus of the ubiquitin moiety in natural or
CC       engineered linear fusion proteins, irrespective of their size or
CC       the presence of an amino-terminal extension to ubiquitin.
```

```
CC   -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC       ubiquitin + a thiol.
CC   -!- SIMILARITY: Belongs to peptidase family C19.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63484; AAA35189.1; -.
DR   EMBL; Z74170; CAA98690.1; -.
DR   PIR; S67665; S67665.
DR   GermOnline; 140364; -.
DR   MEROPS; C19.002; -.
DR   SGD; S0002280; UBP1.
DR   GO; GO:0005737; C:cytoplasm; IC.
DR   GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.
DR   InterPro; IPR001394; Peptidase_C19.
DR   Pfam; PF00443; UCH; 1.
DR   PROSITE; PS00972; UCH_2_1; 1.
DR   PROSITE; PS00973; UCH_2_2; 1.
DR   PROSITE; PS50235; UCH_2_3; 1.
KW   Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT   ACT_SITE    110    110       BY SIMILARITY.
FT   ACT_SITE    689    689       BY SIMILARITY.
FT   ACT_SITE    697    697       BY SIMILARITY.
FT   CONFLICT    418    418       P -> L (IN REF. 2).
SQ   SEQUENCE   809 AA;  92752 MW;  8E3D0B2919E9EC9A CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 809;
  Best Local Similarity   71.4%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | |||:|
Db        655 DEEDVIP 661


RESULT 12
EBN6_EBV
ID   EBN6_EBV       STANDARD;      PRT;   992 AA.
AC   P03204;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   EBNA-6 nuclear protein (EBNA-3C) (EBNA-4B).
GN   BERF3-BERF4.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=88155772; PubMed=2831394;
RA   Petti L., Sample J., Wang F., Kieff E.;
RT   "A fifth Epstein-Barr virus nuclear protein (EBNA3C) is expressed in
RT   latently infected growth-transformed lymphocytes.";
RL   J. Virol. 62:1330-1338(1988).
RN   [3]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=90266473; PubMed=2161150;
RA   Petti L., Sample C., Kieff E.;
RT   "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT   latent infection nuclear proteins.";
RL   Virology 176:563-574(1990).
CC   -!- FUNCTION: INVOLVED IN LATENT CYCLE.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with the nuclear matrix.
CC   -!- SIMILARITY: SOME SIMILARITIES EXIST BETWEEN EBNA 4, 5, AND 6.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01555; CAA24859.1; -.
DR   InterPro; IPR007706; EBNA-3.
DR   Pfam; PF05009; EBNA-3; 1.
KW   Nuclear protein; Repeat.
FT   DOMAIN       74     80       POLY-ARG.
FT   DOMAIN      551    610       10 X 5 AA TANDEM REPEATS.
FT   DOMAIN      741    779       3 X 13 AA TANDEM REPEATS.
SQ   SEQUENCE   992 AA;  109129 MW;  39BEAB9BC515BD84 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 992;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||| ||
Db         89 DTEDNVP 95


RESULT 13
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
```

```
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1224   1226
FT   TURN       1232   1233
FT   TURN       1236   1238
FT   HELIX      1239   1246
FT   TURN       1247   1248
FT   STRAND     1251   1255
FT   HELIX      1258   1271
FT   TURN       1272   1272
FT   STRAND     1277   1280
FT   TURN       1281   1282
FT   STRAND     1283   1285
FT   STRAND     1291   1295
FT   HELIX      1296   1301
FT   TURN       1302   1303
FT   STRAND     1312   1316
FT   TURN       1317   1319
FT   HELIX      1323   1335
FT   TURN       1336   1340
FT   STRAND     1343   1347
FT   TURN       1352   1353
FT   TURN       1360   1361
FT   STRAND     1362   1366
FT   STRAND     1368   1368
FT   STRAND     1373   1375
FT   TURN       1376   1377
FT   STRAND     1378   1380
FT   HELIX      1382   1385
FT   STRAND     1389   1393
FT   HELIX      1397   1409
FT   TURN       1410   1411
FT   STRAND     1414   1417
FT   TURN       1419   1420
FT   STRAND     1432   1436
FT   TURN       1438   1439
FT   STRAND     1450   1453
FT   STRAND     1456   1463
FT   STRAND     1471   1478
FT   STRAND     1480   1480
FT   HELIX      1481   1488
FT   TURN       1489   1490
FT   STRAND     1497   1501
FT   STRAND     1507   1507
FT   STRAND     1511   1511
```

```
FT   HELIX       1514   1527
FT   HELIX       1532   1544
FT   STRAND      1550   1550
FT   HELIX       1555   1564
FT   HELIX       1570   1578
FT   TURN        1579   1580
FT   HELIX       1584   1597
FT   TURN        1598   1598
FT   HELIX       1606   1611
FT   TURN        1614   1618
FT   STRAND      1622   1623
FT   STRAND      1627   1627
FT   STRAND      1635   1636
FT   HELIX       1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 3011;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db       2413 DTEDVV 2418


RESULT 14
POLG_HCVB
ID   POLG_HCVB      STANDARD;      PRT;  3898 AA.
AC   P21530;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Genome polyprotein.
OS   Hog cholera virus (strain Brescia) (Swine fever virus).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Pestivirus.
OX   NCBI_TaxID=11098;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90281581; PubMed=2162104;
RA   Moormann R.J.M., Warmerdam P.A.M., van der Meer B., Schaaper W.M.M.,
RA   Wensvoort G., Hulst M.M.;
RT   "Molecular cloning and nucleotide sequence of hog cholera virus
RT   strain Brescia and mapping of the genomic region encoding envelope
RT   protein E1.";
RL   Virology 177:184-198(1990).
CC   -!- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN
CC       WITH HELICASE AND PROTEASE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: THE GP51-GP54 PROTEIN IS ANCHORED TO THE
CC       VIRAL ENVELOPE.
CC   -!- SIMILARITY: TO BOVINE VIRAL DIARRHEA VIRUS GENOME POLYPROTEIN.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31768; AAA43043.1; -.
DR   PIR; A35317; GNWVHB.
DR   HSSP; P27958; 1A1V.
DR   MEROPS; C53.001; -.
DR   MEROPS; S31.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR008751; Peptidase_C53.
DR   InterPro; IPR000280; Peptidase_S31.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   InterPro; IPR001568; RNase_T2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF05550; Peptidase_C53; 1.
DR   Pfam; PF05578; Peptidase_S31; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   PRINTS; PR00729; CDVENDOPTASE.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00531; RNASE_T2_2; UNKNOWN_1.
KW   Polyprotein; Glycoprotein; Transmembrane; Hydrolase; Serine protease;
KW   Helicase.
FT   CHAIN        ?1   ?267       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN      ?268   ?500       GP42 (E2) (POTENTIAL).
FT   CHAIN      ?501   ?689       GP31 (E3) (POTENTIAL).
FT   CHAIN      ?690  ?1060       GP51-GP54 (ENVELOPE PROTEIN E1).
FT   CHAIN     ?1611  ?2111       P80 (POTENTIAL).
FT   TRANSMEM   1032   1048       POTENTIAL.
FT   ACT_SITE   1658   1658       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1695   1695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1752   1752       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    332    332       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    500    500       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    805    805       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    810    810       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    949    949       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1713   1713       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2134   2134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2217   2217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2419   2419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2494   2494       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2787   2787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2815   2815       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2891   2891       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3103   3103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3211   3211       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3316   3316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3689   3689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3698   3698       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3794   3794       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3898 AA;  438423 MW;  EC6EB207A09D59FD CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 3898;
  Best Local Similarity   83.3%;  Pred. No. 6.9e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||:|
Db        601 TEDVIP 606


RESULT 15
HP1_DROVI
ID   HP1_DROVI      STANDARD;      PRT;   213 AA.
AC   P29227;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Heterochromatin protein 1 (HP1).
GN   SU(VAR)205 OR HP1.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7244;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93096597; PubMed=1461737;
RA   Clark R.F., Elgin S.C.R.;
RT   "Heterochromatin protein 1, a known suppressor of position-effect
RT   variegation, is highly conserved in Drosophila.";
RL   Nucleic Acids Res. 20:6067-6074(1992).
CC   -!- FUNCTION: Structural component of heterochromatin, involved in
CC       gene repression and the modification of position-effect-
CC       variegation.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 2 chromo domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88753; AAB00733.1; -.
DR   PIR; S35522; S35522.
DR   HSSP; P23197; 1AP0.
DR   FlyBase; FBgn0013097; Dvir\Su(var)205.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR008251; Chromo_shadow.
DR   Pfam; PF00385; chromo; 1.
DR   Pfam; PF01393; Chromo_shadow; 1.
DR   PRINTS; PR00504; CHROMODOMAIN.
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00300; ChSh; 1.
DR   PROSITE; PS00598; CHROMO_1; 1.
DR   PROSITE; PS50013; CHROMO_2; 2.
KW   Chromatin regulator; Nuclear protein; Transcription regulation;
KW   Repressor; Repeat.
FT   DOMAIN       24     82       CHROMO 1.
FT   DOMAIN      154    212       CHROMO 2.
FT   DOMAIN       18     23       POLY-GLU.
FT   DOMAIN       83     89       POLY-SER.
SQ   SEQUENCE   213 AA;  23451 MW;  260888892FCFB875 CRC64;

  Query Match             76.3%;  Score 29;  DB 1;  Length 213;
  Best Local Similarity   71.4%;  Pred. No. 43;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              || |:||
Db        138 DTGDIVP 144


Search completed: March 31, 2004, 16:46:22
Job time : 12.4667 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:40:57 ; Search time 34.1333 Seconds
(without alignments)
73.950 Million cell updates/sec

Title: US-09-909-077-4
Perfect score: 38
Sequence: 1 DTEDVVPX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	89.5	941	16	Q81AL6	Q81al6 bacillus ce
2	34	89.5	2358	16	Q81YE8	Q81ye8 bacillus an
3	33	86.8	415	16	Q8FKP6	Q8fkp6 escherichia
4	33	86.8	1208	12	Q8VB56	Q8vb56 white spot
5	33	86.8	1208	12	Q91LF8	Q911f8 white spot
6	32	84.2	134	16	Q8FE11	Q8fe11 escherichia
7	32	84.2	437	10	Q39595	Q39595 chlamydomon
8	32	84.2	580	10	Q9SHK9	Q9shk9 arabidopsis
9	32	84.2	1323	16	Q8ZF67	Q8zf67 yersinia pe
10	32	84.2	1323	16	Q8D0B6	Q8d0b6 yersinia pe
11	31	81.6	290	5	Q9VCD5	Q9vcd5 drosophila
12	31	81.6	324	9	Q38582	Q38582 bacteriopha
13	31	81.6	349	2	Q9WXH3	Q9wxh3 alcaligenes
14	31	81.6	364	16	Q9RJ80	Q9rj80 streptomyce
15	31	81.6	384	10	O64691	O64691 arabidopsis
16	31	81.6	422	12	Q68085	Q68085 classical s
17	31	81.6	547	16	Q8D9H4	Q8d9h4 vibrio vuln
18	31	81.6	637	10	O49524	O49524 arabidopsis
19	31	81.6	642	10	Q8L7S5	Q8l7s5 arabidopsis
20	31	81.6	762	2	Q9LCC0	Q9lcc0 alcaligenes
21	31	81.6	822	12	Q9IBT4	Q9ibt4 turkey herp
22	31	81.6	3898	12	Q918S9	Q918s9 classical s
23	31	81.6	3898	12	Q8UZK1	Q8uzk1 classical s
24	30	78.9	136	12	Q69142	Q69142 human herpe
25	30	78.9	139	5	Q95X76	Q95x76 caenorhabdi
26	30	78.9	156	12	Q8B4N0	Q8b4n0 red sea bre
27	30	78.9	162	11	Q922J8	Q922j8 mus musculu
28	30	78.9	168	5	P91969	P91969 trichinella
29	30	78.9	181	12	Q81730	Q81730 hepatitis c
30	30	78.9	204	16	Q882B5	Q882b5 pseudomonas
31	30	78.9	222	16	Q8EBT2	Q8ebt2 shewanella
32	30	78.9	235	11	Q8BSU3	Q8bsu3 mus musculu
33	30	78.9	273	16	Q8U7T3	Q8u7t3 agrobacteri
34	30	78.9	281	2	Q8GLU1	Q8glu1 aeromonas j
35	30	78.9	331	16	Q9HYY6	Q9hyy6 pseudomonas
36	30	78.9	335	2	Q9F749	Q9f749 bacteroides
37	30	78.9	339	2	Q9XDJ2	Q9xdj2 bacteroides
38	30	78.9	344	16	Q8PDF0	Q8pdf0 xanthomonas
39	30	78.9	377	10	Q8H6A3	Q8h6a3 stevia reba
40	30	78.9	387	5	O44187	O44187 caenorhabdi
41	30	78.9	392	10	Q84Q44	Q84q44 oryza sativ
42	30	78.9	409	16	Q7UNR5	Q7unr5 rhodopirell
43	30	78.9	412	5	O01819	O01819 caenorhabdi
44	30	78.9	424	6	Q864C2	Q864c2 macaca fasc
45	30	78.9	427	16	Q93HF6	Q93hf6 streptomyce

ALIGNMENTS

RESULT 1
Q81AL6
ID Q81AL6 PRELIMINARY; PRT; 941 AA.
AC Q81AL6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell surface protein.
GN BC3547.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
DR EMBL; AE017009; AAP10481.1; -.
DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
DR InterPro; IPR001434; DUF11.
DR Pfam; PF01345; DUF11; 5.
DR TIGRFAMs; TIGR01451; B_ant_repeat; 5.
KW Complete proteome.
SQ SEQUENCE 941 AA; 97858 MW; D4A2784D95911BAD CRC64;

Query Match 89.5%; Score 34; DB 16; Length 941;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVP 7
              |||||:|
Db         92 DTEDVLP 98
```

```
RESULT 2
Q81YE8
ID   Q81YE8      PRELIMINARY;      PRT;  2358 AA.
AC   Q81YE8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Conserved repeat domain protein.
GN   BA3601.
OS   Bacillus anthracis (strain Ames).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017035; AAP27357.1; -.
DR   TIGR; BA3601; -.
DR   GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR001434; DUF11.
DR   Pfam; PF01345; DUF11; 12.
DR   TIGRFAMs; TIGR01451; B_ant_repeat; 15.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   2358 AA;  245092 MW;  EB8F7457CB7B7312 CRC64;

  Query Match             89.5%;  Score 34;  DB 16;  Length 2358;
  Best Local Similarity   85.7%;  Pred. No. 4e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||||:|
Db         92 DTEDVLP 98


RESULT 3
Q8FKP6
ID   Q8FKP6      PRELIMINARY;      PRT;   415 AA.
AC   Q8FKP6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   C0360.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016756; AAN78841.1; -.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003423; OEP.
DR   Pfam; PF02321; OEP; 2.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   415 AA;  46967 MW;  C28B3CCF9A935959 CRC64;

  Query Match             86.8%;  Score 33;  DB 16;  Length 415;
  Best Local Similarity   71.4%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              ||||::|
Db         79 DTEDILP 85


RESULT 4
Q8VB56
ID   Q8VB56      PRELIMINARY;      PRT;  1208 AA.
AC   Q8VB56;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Wsv139 (WSSV194).
OS   White spot syndrome virus (WSSV).
OC   Viruses; dsDNA viruses, no RNA stage; Nimaviridae.
OX   NCBI_TaxID=92652;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21548311; PubMed=11689662;
RA   Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;
RT   "Complete genome sequence of the shrimp white spot bacilliform
RT   virus.";
RL   J. Virol. 75:11811-11820(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Taiwan;
RX   MEDLINE=20517548; PubMed=11062040;
RA   Tsai M.F., Yu H.T., Tzeng H.F., Leu J.H., Chou C.M., Huang C.J.,
RA   Wang C.H., Lin J.Y., Kou G.H., Lo C.F.;
RT   "Identification and characterization of a shrimp white spot syndrome
RT   virus (WSSV) gene that encodes a novel chimeric polypeptide of
RT   cellular-type thymidine kinase and thymidylate kinase.";
RL   Virology 277:100-110(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Taiwan;
RX   MEDLINE=21844071; PubMed=11853398;
RA   Chen L.L., Leu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,
RA   Lo C.F., Kou G.H.;
RT   "Identification of a nucleocapsid protein (VP35) gene of shrimp white
RT   spot syndrome virus and characterization of the motif important for
RT   targeting VP35 to the nuclei of transfected insect cells.";
RL   Virology 293:44-53(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Taiwan;
RA   Lo C.-F., Kou G.-H.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF332093; AAL33143.1; -.
DR   EMBL; AF440570; AAL89062.1; -.
SQ   SEQUENCE   1208 AA;  137842 MW;  F23A00D0D46F33EF CRC64;
```

```
  Query Match             86.8%;  Score 33;  DB 12;  Length 1208;
  Best Local Similarity   71.4%;  Pred. No. 3.3e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||::|
Db        747 DTEDLIP 753


RESULT 5
Q91LF8
ID   Q91LF8      PRELIMINARY;      PRT;  1208 AA.
AC   Q91LF8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF82.
OS   White spot syndrome virus (WSSV).
OC   Viruses; dsDNA viruses, no RNA stage; Nimaviridae.
OX   NCBI_TaxID=92652;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21342572; PubMed=11448154;
RA   van Hulten M.C.W., Witteveldt J., Peters S., Kloosterboer N.,
RA   Tarchini R., Fiers M., Sandbrink H., Lankhorst R.K., Vlak J.M.;
RT   "The white spot syndrome virus DNA genome sequence.";
RL   Virology 286:7-22(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   van Hulten M.C.W., Witteveldt J., Peters S., Kloosterboer N.,
RA   Tarchini R., Fiers M., Sandbrink H., Lankhorst R.K., Vlak J.M.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369029; AAK77751.1; -.
SQ   SEQUENCE   1208 AA;  137793 MW;  0D1BC0C55E3D2CEF CRC64;

  Query Match             86.8%;  Score 33;  DB 12;  Length 1208;
  Best Local Similarity   71.4%;  Pred. No. 3.3e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||::|
Db        747 DTEDLIP 753


RESULT 6
Q8FE11
ID   Q8FE11      PRELIMINARY;      PRT;   134 AA.
AC   Q8FE11;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   C3560.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016766; AAN82008.1; -.
DR   InterPro; IPR002514; Transposase_8.
DR   Pfam; PF01527; Transposase_8; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   134 AA;  15224 MW;  E3C417AD72CC92E3 CRC64;

  Query Match             84.2%;  Score 32;  DB 16;  Length 134;
  Best Local Similarity   85.7%;  Pred. No. 53;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | |||||
Db         78 DAEDVVP 84


RESULT 7
Q39595
ID   Q39595      PRELIMINARY;      PRT;   437 AA.
AC   Q39595;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   ATP sulfurylase Ats1.
GN   ATS1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CC125;
RA   Yildiz F.H, Davies J.P, Grossman A.;
RT   "Controlled expression of the ATP sulfurylase gene of Chlamydomonas
RT   reinhardtii.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U57088; AAB01234.1; -.
DR   PIR; T08094; T08094.
DR   GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA.
DR   GO; GO:0000103; P:sulfate assimilation; IEA.
DR   InterPro; IPR002650; ATP-sulfurylase.
DR   Pfam; PF01747; ATP-sulfurylase; 1.
DR   ProDom; PD002381; ATP-sulfurylase; 1.
DR   TIGRFAMs; TIGR00339; sopT; 1.
SQ   SEQUENCE   437 AA;  47747 MW;  D22013FBD471AF6E CRC64;

  Query Match             84.2%;  Score 32;  DB 10;  Length 437;
  Best Local Similarity   71.4%;  Pred. No. 1.8e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |:||:||
Db        138 DSEDIVP 144


RESULT 8
Q9SHK9
ID   Q9SHK9      PRELIMINARY;      PRT;   580 AA.
AC   Q9SHK9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   At2g17600 protein.
GN   AT2G17600.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
```

```
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007584; AAD32917.1; -.
DR   PIR; B84554; B84554.
DR   InterPro; IPR004146; DC1.
DR   Pfam; PF03107; DC1; 5.
SQ   SEQUENCE   580 AA;  67315 MW;  0151E70CF7F6BB20 CRC64;

  Query Match             84.2%;  Score 32;  DB 10;  Length 580;
  Best Local Similarity   71.4%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              :|||:||
Db        266 ETEDIVP 272


RESULT 9
Q8ZF67
ID   Q8ZF67      PRELIMINARY;      PRT;  1323 AA.
AC   Q8ZF67;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Bifunctional PutA protein (EC 1.5.99.8).
GN   PUTA OR YPO1851.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414150; CAC90668.1; -.
DR   PIR; AH0225; AH0225.
DR   GO; GO:0003842; F:1-pyrroline-5-carboxylate dehydrogenase act. . .; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0004657; F:proline dehydrogenase activity; IEA.
DR   GO; GO:0006537; P:glutamate biosynthesis; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   GO; GO:0006561; P:proline biosynthesis; IEA.
DR   GO; GO:0006562; P:proline catabolism; IEA.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   InterPro; IPR005933; D1pyr5carbox3.
DR   InterPro; IPR002872; Pro_dh.
DR   Pfam; PF00171; aldedh; 1.
DR   Pfam; PF01619; Pro_dh; 1.
DR   TIGRFAMs; TIGR01238; D1pyr5carbox3; 1.
DR   PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR   PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW   Oxidoreductase; Complete proteome.
SQ   SEQUENCE   1323 AA;  144380 MW;  291B1B54136C7FE2 CRC64;

  Query Match             84.2%;  Score 32;  DB 16;  Length 1323;
  Best Local Similarity   71.4%;  Pred. No. 5.9e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||| :|
Db         63 DTEDAIP 69


RESULT 10
Q8D0B6
ID   Q8D0B6      PRELIMINARY;      PRT;  1323 AA.
AC   Q8D0B6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Proline dehydrogenase, P5C dehydrogenase.
GN   PUTA OR Y2455.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
DR   EMBL; AE013849; AAM86012.1; -.
DR   GO; GO:0003842; F:1-pyrroline-5-carboxylate dehydrogenase act. . .; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0004657; F:proline dehydrogenase activity; IEA.
DR   GO; GO:0006537; P:glutamate biosynthesis; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   GO; GO:0006561; P:proline biosynthesis; IEA.
DR   GO; GO:0006562; P:proline catabolism; IEA.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   InterPro; IPR005933; D1pyr5carbox3.
DR   InterPro; IPR002872; Pro_dh.
DR   Pfam; PF00171; aldedh; 1.
DR   Pfam; PF01619; Pro_dh; 1.
DR   TIGRFAMs; TIGR01238; D1pyr5carbox3; 1.
DR   PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR   PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
SQ   SEQUENCE   1323 AA;  144353 MW;  AA69CB26C099AF2D CRC64;

  Query Match             84.2%;  Score 32;  DB 16;  Length 1323;
  Best Local Similarity   71.4%;  Pred. No. 5.9e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              |||| :|
Db         63 DTEDAIP 69


RESULT 11
Q9VCD5
ID   Q9VCD5      PRELIMINARY;      PRT;   290 AA.
AC   Q9VCD5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG10694 protein (AT15685p).
GN   CG10694.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003746; AAF56234.1; -.
DR   EMBL; AY089339; AAL90077.1; -.
DR   HSSP; P54725; 1DV0.
DR   FlyBase; FBgn0039147; CG10694.
DR   InterPro; IPR000449; UBA_domain.
DR   InterPro; IPR000626; Ubiquitin.
DR   Pfam; PF00627; UBA; 2.
DR   Pfam; PF00240; ubiquitin; 1.
DR   SMART; SM00165; UBA; 2.
DR   SMART; SM00213; UBQ; 1.
DR   PROSITE; PS50030; UBA; 2.
DR   PROSITE; PS50053; UBIQUITIN_2; 1.
SQ   SEQUENCE   290 AA;  32557 MW;  2EBE4FAAFE354DE2 CRC64;

  Query Match             81.6%;  Score 31;  DB 5;  Length 290;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||||
Db        101 TEDVVP 106


RESULT 12
Q38582
ID   Q38582      PRELIMINARY;      PRT;   324 AA.
AC   Q38582;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Coat protein.
GN   13.
OS   Bacteriophage SPP1.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   Lambda-like viruses.
OX   NCBI_TaxID=10724;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Becker B., Gassel M., Tavares P., Lurz R., Alonso J.C.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Alonso J.C.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X89721; CAA61870.1; -.
DR   EMBL; X97918; CAA66544.1; -.
DR   PIR; S58142; S58142.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
KW   Coat protein.
SQ   SEQUENCE   324 AA;  35354 MW;  E3BD2B06D702160D CRC64;

  Query Match             81.6%;  Score 31;  DB 9;  Length 324;
  Best Local Similarity   71.4%;  Pred. No. 2.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
                |:|:||
Db         71 DTDDLVP 77


RESULT 13
Q9WXH3
ID   Q9WXH3      PRELIMINARY;      PRT;   349 AA.
AC   Q9WXH3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   1-hydroxy-2-naphthoate dioxygenase.
GN   PHNG.
OS   Alcaligenes faecalis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Alcaligenes.
OX   NCBI_TaxID=511;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AFK2;
RA   Kiyohara H., Tabata Y., Takizawa N.;
RT   "A phenanthrene degradative gene cluster in Alcaligenes faecalis
RT   AFK2.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB024945; BAA76328.1; -.
DR   GO; GO:0016702; F:oxidoreductase activity, acting on single d. . .; IEA.
DR   InterPro; IPR007113; Cupin_sup.
KW   Dioxygenase.
SQ   SEQUENCE   349 AA;  38706 MW;  8C311EB2AE68DA49 CRC64;

  Query Match             81.6%;  Score 31;  DB 2;  Length 349;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TEDVVP 7
              ||||||
```

```
Db         11 TEDVVP 16


RESULT 14
Q9RJ80
ID   Q9RJ80      PRELIMINARY;      PRT;   364 AA.
AC   Q9RJ80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glycine betaine transport ATP-binding protein.
GN   OPUAA OR SCO1621 OR SCI41.04C.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AL939109; CAB59474.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0015171; F:amino acid transporter activity; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006865; P:amino acid transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR000644; CBS_domain.
DR   InterPro; IPR005892; ProV.
DR   Pfam; PF00005; ABC_tran; 1.
DR   Pfam; PF00571; CBS; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00116; CBS; 1.
DR   TIGRFAMs; TIGR01186; proV; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW   ATP-binding; Transport; Complete proteome.
SQ   SEQUENCE   364 AA;  39970 MW;  B67DE74839610491 CRC64;

  Query Match             81.6%;  Score 31;  DB 16;  Length 364;
  Best Local Similarity   85.7%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | |||||
Db        340 DDEDVVP 346


RESULT 15
O64691
ID   O64691      PRELIMINARY;      PRT;   384 AA.
AC   O64691;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative katanin.
GN   AT2G34560.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA   Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA   Somerville C.R., Venter J.C.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Town C.D., Kaul S.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004077; AAC26698.2; -.
DR   EMBL; AY084858; AAM61422.1; -.
DR   PIR; B84758; B84758.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   Pfam; PF00004; AAA; 1.
SQ   SEQUENCE   384 AA;  42872 MW;  ACCF8C8726ABD386 CRC64;

  Query Match             81.6%;  Score 31;  DB 10;  Length 384;
  Best Local Similarity   85.7%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVP 7
              | |||||
Db        332 DREDVVP 338


Search completed: March 31, 2004, 16:48:39
Job time : 36.1333 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:40:57 ; Search time 34.1333 Seconds
(without alignments)
73.950 Million cell updates/sec

Title: US-09-909-077-1
Perfect score: 32
Sequence: 1 DTEDVVXX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	93.8	181	12	Q81730	Q81730 hepatitis c
2	30	93.8	235	11	Q8BSU3	Q8bsu3 mus musculu
3	30	93.8	344	16	Q8PDF0	Q8pdf0 xanthomonas
4	30	93.8	412	5	O01819	O01819 caenorhabdi
5	30	93.8	424	6	Q864C2	Q864c2 macaca fasc
6	30	93.8	439	5	O44918	O44918 caenorhabdi
7	30	93.8	448	12	Q9E3A2	Q9e3a2 hepatitis c
8	30	93.8	448	12	Q9E364	Q9e364 hepatitis c
9	30	93.8	448	12	Q9E3D6	Q9e3d6 hepatitis c
10	30	93.8	448	12	Q9E3A6	Q9e3a6 hepatitis c
11	30	93.8	448	12	Q9E388	Q9e388 hepatitis c
12	30	93.8	448	12	Q9E363	Q9e363 hepatitis c
13	30	93.8	448	12	Q9E3D8	Q9e3d8 hepatitis c
14	30	93.8	448	12	Q9E3A7	Q9e3a7 hepatitis c
15	30	93.8	448	12	Q9E3A4	Q9e3a4 hepatitis c
16	30	93.8	448	12	Q9E3D2	Q9e3d2 hepatitis c
17	30	93.8	448	12	Q9E385	Q9e385 hepatitis c
18	30	93.8	448	12	Q9E398	Q9e398 hepatitis c
19	30	93.8	448	12	Q9E3D4	Q9e3d4 hepatitis c
20	30	93.8	448	12	Q9E381	Q9e381 hepatitis c
21	30	93.8	448	12	Q9E392	Q9e392 hepatitis c
22	30	93.8	448	12	Q9E3A8	Q9e3a8 hepatitis c
23	30	93.8	448	12	Q9E390	Q9e390 hepatitis c
24	30	93.8	448	12	Q9E3D7	Q9e3d7 hepatitis c
25	30	93.8	448	12	Q9E3A1	Q9e3a1 hepatitis c
26	30	93.8	448	12	Q9E387	Q9e387 hepatitis c
27	30	93.8	448	12	Q9E3B0	Q9e3b0 hepatitis c
28	30	93.8	448	12	Q9E3E2	Q9e3e2 hepatitis c
29	30	93.8	448	12	Q9E391	Q9e391 hepatitis c
30	30	93.8	448	12	Q9E3A9	Q9e3a9 hepatitis c
31	30	93.8	448	12	Q9E3B1	Q9e3b1 hepatitis c
32	30	93.8	448	12	Q9E394	Q9e394 hepatitis c
33	30	93.8	448	12	Q9E365	Q9e365 hepatitis c
34	30	93.8	448	12	Q9E3C6	Q9e3c6 hepatitis c
35	30	93.8	448	12	Q9E369	Q9e369 hepatitis c
36	30	93.8	448	12	Q9E3A3	Q9e3a3 hepatitis c
37	30	93.8	448	12	Q9E3E0	Q9e3e0 hepatitis c
38	30	93.8	448	12	Q9E370	Q9e370 hepatitis c
39	30	93.8	448	12	Q9E384	Q9e384 hepatitis c
40	30	93.8	448	12	Q9E3C8	Q9e3c8 hepatitis c
41	30	93.8	448	12	Q9E3D3	Q9e3d3 hepatitis c
42	30	93.8	448	12	Q9E3A5	Q9e3a5 hepatitis c
43	30	93.8	448	12	Q9E383	Q9e383 hepatitis c
44	30	93.8	448	12	Q9E397	Q9e397 hepatitis c
45	30	93.8	448	12	Q9E3C9	Q9e3c9 hepatitis c

ALIGNMENTS

RESULT 1
Q81730
ID Q81730 PRELIMINARY; PRT; 181 AA.
AC Q81730;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Potential NS5 domain; putative (Genome polyprotein)
DE (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RA Inchauspe G., Zebedee S.L, Nasoff M.S., Sugitani M., Abe K.,
RA Prince A.M.;
RT "Cloning and nucleotide sequence analysis of structural and
RT nonstructural regions of the hutchinson strain of hepatitus C.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Hutchinson;
RX MEDLINE=89222455; PubMed=2496467;
RA Kuo G., Choo Q.-L., Alter H.J., Gitnick G.L., Redeker A.G.,
RA Purcell R.H., Miyamura T., Dienstag J.L., Alter M.J., Stevens C.E.,
RA Tegtmeier G.E., Bonino F., Colombo M., Lee W.-S., Kuo C., Berger K.,
RA Shuster J.R., Overby L.R., Bradley D.W., Houghton M.;
RT "An assay for circulating antibodies to a major etiologic virus of
RT human non-A, non-B hepatitis.";

```
RL   Science 244:362-365(1992).
DR   EMBL; M55974; AAA45663.1; -.
DR   GO; GO:0019012; C:virion; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   InterPro; IPR002166; HCV_RdRP.
DR   Pfam; PF00998; Viral_RdRP; 1.
KW   Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW   Transferase.
FT   NON_TER       1      1
FT   NON_TER     181    181
SQ   SEQUENCE   181 AA;  19124 MW;  7C30235E19009C4A CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         98 DTEDVV 103


RESULT 2
Q8BSU3
ID   Q8BSU3      PRELIMINARY;      PRT;   235 AA.
AC   Q8BSU3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Pituitary;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK030512; BAC26998.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   235 AA;  25133 MW;  50928397A1AD126F CRC64;

  Query Match             93.8%;  Score 30;  DB 11;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         92 DTEDVV 97


RESULT 3
Q8PDF0
ID   Q8PDF0      PRELIMINARY;      PRT;   344 AA.
AC   Q8PDF0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Biotin synthase.
GN   BIOB OR XCC0388.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE012135; AAM39707.1; -.
DR   GO; GO:0004076; F:biotin synthase activity; IEA.
DR   GO; GO:0009102; P:biotin biosynthesis; IEA.
DR   InterPro; IPR002684; Biotin_synth.
DR   InterPro; IPR007197; Radical_SAM.
DR   Pfam; PF04055; Radical_SAM; 1.
DR   TIGRFAMs; TIGR00433; bioB; 1.
KW   Complete proteome.
SQ   SEQUENCE   344 AA;  37515 MW;  4C576AF23C7F7D76 CRC64;

  Query Match             93.8%;  Score 30;  DB 16;  Length 344;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         79 DTEDVV 84


RESULT 4
O01819
ID   O01819      PRELIMINARY;      PRT;   412 AA.
AC   O01819;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F57C9.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Geisel C., Kramer J., Gibson A.;
RT   "The sequence of C. elegans cosmid F57C9.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
```

```
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF003142; AAB54190.1; -.
DR   PIR; T15214; T15214.
DR   WormPep; F57C9.7; CE11334.
KW   Hypothetical protein.
SQ   SEQUENCE   412 AA;  47344 MW;  444930E65B999C71 CRC64;

  Query Match             93.8%;  Score 30;  DB 5;  Length 412;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        132 DTEDVV 137


RESULT 5
Q864C2
ID   Q864C2      PRELIMINARY;      PRT;   424 AA.
AC   Q864C2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ZPC protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris J.D., Piersen C.E.;
RT   "Cloning and Expression of Cynomolgus Monkey and Baboon Zona Pellucida
RT   Proteins.";
RL   Mol. Reprod. Dev. 0:0-0(2003).
DR   EMBL; AY222644; AAP13258.1; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
SQ   SEQUENCE   424 AA;  46997 MW;  7DD6784E75280B3B CRC64;

  Query Match             93.8%;  Score 30;  DB 6;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 6
O44918
ID   O44918      PRELIMINARY;      PRT;   439 AA.
AC   O44918;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   W10G11.17.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Goela D., Scheet P.;
RT   "The sequence of C. elegans cosmid W10G11.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF040661; AAG24223.1; -.
DR   PIR; G88103; G88103.
DR   WormPep; W10G11.17; CE14832.
KW   Hypothetical protein.
SQ   SEQUENCE   439 AA;  49291 MW;  045B32C0804DEB0E CRC64;

  Query Match             93.8%;  Score 30;  DB 5;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         40 DTEDVV 45


RESULT 7
Q9E3A2
ID   Q9E3A2      PRELIMINARY;      PRT;   448 AA.
AC   Q9E3A2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265037; AAG21176.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48876 MW;  5F15BF6005A08CBE CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 8
Q9E364
ID   Q9E364      PRELIMINARY;      PRT;   448 AA.
AC   Q9E364;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265075; AAG21214.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48747 MW;  123D101C4612F1CB CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 9
Q9E3D6
ID   Q9E3D6      PRELIMINARY;      PRT;   448 AA.
AC   Q9E3D6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265002; AAG21142.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48945 MW;  2EF1E1EE92C22658 CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 10
Q9E3A6
ID   Q9E3A6      PRELIMINARY;      PRT;   448 AA.
AC   Q9E3A6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265032; AAG21172.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48777 MW;  0728101C460841CB CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 11
Q9E388
ID   Q9E388      PRELIMINARY;      PRT;   448 AA.
AC   Q9E388;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265051; AAG21190.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48899 MW;  2A978EE71887B459 CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 12
Q9E363
ID   Q9E363      PRELIMINARY;      PRT;   448 AA.
AC   Q9E363;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265076; AAG21215.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48807 MW;  072810055F0858D2 CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 13
Q9E3D8
ID   Q9E3D8      PRELIMINARY;      PRT;   448 AA.
AC   Q9E3D8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265000; AAG21140.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48943 MW;  2E8F9EDB92E6BFAF CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 14
Q9E3A7
ID   Q9E3A7      PRELIMINARY;      PRT;   448 AA.
AC   Q9E3A7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265031; AAG21171.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48777 MW;  0728101C460841CB CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


RESULT 15
Q9E3A4
ID   Q9E3A4      PRELIMINARY;      PRT;   448 AA.
AC   Q9E3A4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyprotein (Fragment).
OS   Hepatitis C virus type 1a.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=31646;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20438098; PubMed=10982347;
RA   Nousbaum J., Polyak S.J., Ray S.C., Sullivan D.G., Larson A.M.,
RA   Carithers R.L. Jr., Gretch D.R.;
RT   "Prospective characterization of full-length hepatitis C virus NS5A
RT   quasispecies during induction and combination antiviral therapy.";
RL   J. Virol. 74:9028-9038(2000).
DR   EMBL; AF265035; AAG21174.1; -.
DR   InterPro; IPR002868; HCV_NS5a.
DR   Pfam; PF01506; HCV_NS5a; 1.
FT   NON_TER       1      1
FT   NON_TER     448    448
SQ   SEQUENCE   448 AA;  48777 MW;  0728101C460841CB CRC64;

  Query Match             93.8%;  Score 30;  DB 12;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        441 DTEDVV 446


Search completed: March 31, 2004, 16:48:33
Job time : 36.1333 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:01 ; Search time 38 Seconds
(without alignments)
44.613 Million cell updates/sec

Title: US-09-909-077-2
Perfect score: 25
Sequence: 1 DXLIXC 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	92.0	105	7	ADE08222	Ade08222 Novel pro
2	23	92.0	120	5	ABP63947	Abp63947 Human ORF
3	23	92.0	206	4	AAU62959	Aau62959 Propionib
4	23	92.0	206	6	ABM59478	Abm59478 Propionib
5	23	92.0	216	4	ABG10921	Abg10921 Novel hum
6	23	92.0	310	4	AAM41607	Aam41607 Human pol
7	23	92.0	325	7	ADE09054	Ade09054 Novel pro
8	23	92.0	351	6	ABJ19522	Abj19522 UL15 DNA
9	23	92.0	351	6	ABJ19477	Abj19477 UL15 DNA
10	23	92.0	409	6	ABU41410	Abu41410 Protein e
11	23	92.0	616	6	ABM68320	Abm68320 Photorhab
12	23	92.0	628	6	ADA55635	Ada55635 Human pro
13	23	92.0	665	6	ABP98882	Abp98882 Human mol
14	23	92.0	682	6	ABJ19478	Abj19478 UL15 DNA
15	23	92.0	683	6	ABJ19521	Abj19521 UL15 DNA
16	23	92.0	683	6	ABJ19523	Abj19523 UL15 DNA
17	23	92.0	683	6	ABJ19476	Abj19476 UL15 DNA
18	23	92.0	688	7	ADD25205	Add25205 Fertility
19	23	92.0	769	4	ABB65620	Abb65620 Drosophil
20	23	92.0	1207	4	ABB62248	Abb62248 Drosophil
21	23	92.0	1291	4	ABB62147	Abb62147 Drosophil
22	23	92.0	1291	4	ABB66058	Abb66058 Drosophil
23	23	92.0	1569	4	ABG15232	Abg15232 Novel hum
24	23	92.0	2144	4	AAB85029	Aab85029 Protein e
25	22	88.0	18	3	AAB09263	Aab09263 Hepatitis
26	22	88.0	25	7	ADC99927	Adc99927 Murine Me
27	22	88.0	29	4	AAM18817	Aam18817 Peptide #
28	22	88.0	29	4	ABB37923	Abb37923 Peptide #
29	22	88.0	29	4	AAM31333	Aam31333 Peptide #
30	22	88.0	29	4	ABB23178	Abb23178 Protein #
31	22	88.0	29	4	AAM71056	Aam71056 Human bon
32	22	88.0	29	4	AAM58555	Aam58555 Human bra
33	22	88.0	29	4	ABG52771	Abg52771 Human liv
34	22	88.0	29	5	ABG40849	Abg40849 Human pep
35	22	88.0	49	4	AAU27364	Aau27364 Novel bon
36	22	88.0	59	4	AAU42854	Aau42854 Propionib
37	22	88.0	59	6	ABM39373	Abm39373 Propionib
38	22	88.0	98	7	ADC00353	Adc00353 Enterohae
39	22	88.0	107	7	ADA49409	Ada49409 Multi-epi
40	22	88.0	108	2	AAY37690	Aay37690 Chlamydia
41	22	88.0	126	4	ABB64514	Abb64514 Drosophil
42	22	88.0	126	4	ABB65842	Abb65842 Drosophil
43	22	88.0	135	3	AAG00793	Aag00793 Human sec
44	22	88.0	143	3	AAB40841	Aab40841 Human ORF
45	22	88.0	143	5	ABP07709	Abp07709 Human ORF

ALIGNMENTS

RESULT 1
ADE08222
ID ADE08222 standard; protein; 105 AA.
XX
AC ADE08222;
XX
DT 29-JAN-2004 (first entry)
XX
DE Novel protein (useful for identifying genetic disorders) #377.
XX
KW novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder.
XX
OS Unidentified.
XX
PN WO2003054152-A2.
XX
PD 03-JUL-2003.
XX
PF 10-DEC-2002; 2002WO-US039555.
XX
PR 10-DEC-2001; 2001US-0339739P.
PR 11-DEC-2001; 2001US-0339453P.
PR 14-MAR-2002; 2002US-0365091P.
PR 14-MAR-2002; 2002US-0365384P.
PR 12-APR-2002; 2002US-0372381P.
PR 12-APR-2002; 2002US-0372615P.
PR 22-APR-2002; 2002US-00128558.
PR 24-APR-2002; 2002US-0376045P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX
DR WPI; 2003-569235/53.
DR N-PSDB; ADE07311.
XX
PT New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
PS Claim 20; SEQ ID NO 1288; 1177pp; English.
XX
CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially

CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence represents a protein
CC of the invention.
XX
SQ Sequence 105 AA;

Query Match 92.0%; Score 23; DB 7; Length 105;
Best Local Similarity 66.7%; Pred. No. 6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| || |
Db 24 DTLIAC 29

RESULT 2
ABP63947
ID ABP63947 standard; protein; 120 AA.
XX
AC ABP63947;
XX
DT 04-NOV-2002 (first entry)
XX
DE Human ORF317.
XX
KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW cancer; cardiovascular disease; allergy; autoimmune disease;
KW wound healing; blood coagulation disorder; inflammatory disorder.
XX
OS Homo sapiens.
XX
PN US2002082206-A1.
XX
PD 27-JUN-2002.
XX
PF 30-MAY-2001; 2001US-00867550.
XX
PR 30-MAY-2000; 2000US-0208427P.
XX
PA (LEAC/) LEACH M D.
PA (MEHR/) MEHRABAN F.
PA (CONL/) CONLEY P B.
PA (TOPP/) TOPPER J N.
PA (LAWD/) LAW D.
XX
PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
XX
DR WPI; 2002-626554/67.
DR N-PSDB; ABQ98510.
XX
PT New polypeptide designated ORFX are present in human atherogenic cells
PT and are useful to prevent and treat ORFX-associated disorders including
PT cancer, allergy, wound healing or autoimmune, cardiovascular or
PT inflammatory disease.
XX
PS Claim 10; SEQ ID NO 634; 78pp; English.
XX
CC The present invention relates to novel human ORFX polypeptides and their
CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
CC were discovered in human atherogenic cells, in particular in platelets
CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
CC many other tissues as well. Atherogenic cells are cells which have the
CC potential to develop atherosclerotic plaques. The ORFX polypeptides and
CC nucleic acids are useful for treating or preventing a pathological
CC condition associated with an ORFX-associated disorder, e.g. cancer,
CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
CC coagulation disorders or inflammatory disorders. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the USPTO web site at
CC seqdata.uspto.gov/sequence.html?DocID=20020082206
XX
SQ Sequence 120 AA;

Query Match 92.0%; Score 23; DB 5; Length 120;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| || |
Db 17 DALIAC 22

RESULT 3
AAU62959
ID AAU62959 standard; protein; 206 AA.
XX
AC AAU62959;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #23855.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59630.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 24154; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 206 AA;

Query Match 92.0%; Score 23; DB 4; Length 206;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| || |
Db 43 DALIAC 48

RESULT 4
ABM59478
ID ABM59478 standard; protein; 206 AA.
XX
AC ABM59478;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #24154.
XX
KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglass J;
XX
DR WPI; 2003-381789/36.
DR N-PSDB; ACF64559.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Example 1; SEQ ID NO 24154; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an CRF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 206 AA;

Query Match 92.0%; Score 23; DB 6; Length 206;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| || |
Db 43 DALIAC 48

RESULT 5
ABG10921
ID ABG10921 standard; protein; 216 AA.
XX
AC ABG10921;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10912.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS75108.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 41280; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this

```
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 216 AA;

  Query Match             92.0%;  Score 23;  DB 4;  Length 216;
  Best Local Similarity   66.7%;  Pred. No. 1.2e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        199 DSLISC 204


RESULT 6
AAM41607
ID   AAM41607 standard; protein; 310 AA.
XX
AC   AAM41607;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 6538.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J,  Zhao QA;
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60763.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Example 2; SEQ ID NO 6538; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 310 AA;

  Query Match             92.0%;  Score 23;  DB 4;  Length 310;
  Best Local Similarity   66.7%;  Pred. No. 1.6e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db          8 DALITC 13


RESULT 7
ADE09054
ID   ADE09054 standard; protein; 325 AA.
XX
AC   ADE09054;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Novel protein-related contig polypeptide sequence #120.
XX
KW   novel gene; novel protein; tissue marker; molecular weight marker;
KW   chromosome marker; genetic disorder; contig.
XX
OS   Unidentified.
XX
PN   WO2003054152-A2.
XX
PD   03-JUL-2003.
XX
PF   10-DEC-2002; 2002WO-US039555.
XX
PR   10-DEC-2001; 2001US-0339739P.
PR   11-DEC-2001; 2001US-0339453P.
PR   14-MAR-2002; 2002US-0365091P.
PR   14-MAR-2002; 2002US-0365384P.
PR   12-APR-2002; 2002US-0372381P.
PR   12-APR-2002; 2002US-0372615P.
PR   22-APR-2002; 2002US-00128558.
PR   24-APR-2002; 2002US-0376045P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Asundi V,  Goodrich RW,  Ren F,  Zhang J,  Zhao QA,  Wang J;
PI   Ghosh M,  Xue AJ,  Wehrman T,  Weng G,  Zhou P,  Drmanac RT,  Wang Z;
PI   Ma Y,  Wang D,  Chen R,  Xu C,  Boyle BJ;
XX
DR   WPI; 2003-569235/53.
XX
PT   New polynucleotides, useful for expressing recombinant proteins for
PT   analysis, characterization or therapeutic use, or as markers for tissues
PT   in which the corresponding protein is preferentially expressed.
XX
PS   Disclosure; SEQ ID NO 2598; 1177pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of novel
CC   proteins. The DNA and protein sequences of the invention are useful as:
CC   markers for tissues in which the corresponding protein is preferentially
CC   expressed; as molecular weight markers on gels; as chromosome markers or
CC   tags; to identify chromosomes or to map related gene positions; and to
CC   compare with endogenous DNA sequences in patients to identify potential
CC   genetic disorders. The present amino acid sequence was used in the
CC   exemplification of the invention.
XX
SQ   Sequence 325 AA;
```

Query Match 92.0%; Score 23; DB 7; Length 325;
Best Local Similarity 66.7%; Pred. No. 1.7e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db        130 DSLIAC 135
```

RESULT 8
ABJ19522
ID ABJ19522 standard; protein; 351 AA.
XX
AC ABJ19522;
XX
DT 27-MAR-2003 (first entry)
XX
DE UL15 DNA packaging protein #24.
XX
KW Panel; degenerate primer pair; screening; virus family;
KW high throughput screening; UL15 DNA packaging protein; VIDA.
XX
OS Unidentified.
XX
PN WO200299130-A2.
XX
PD 12-DEC-2002.
XX
PF 07-JUN-2002; 2002WO-GB002642.
XX
PR 07-JUN-2001; 2001GB-00013907.
XX
PA (UNLO) UNIV COLLEGE LONDON.
XX
PI Griffiths DJ, Kellam P, Weiss RA;
XX
DR WPI; 2003-148677/14.
XX
PT High throughput method using degenerate polymerase chain reaction primers
PT useful for screening a biological sample for unknown viruses.
XX
PS Disclosure; Page 24; 39pp; English.
XX
CC The invention relates to a novel method for designing a panel of
CC degenerate primer pairs for screening for new members of multiple known
CC virus families in a biological sample. The method can be used in high
CC throughput screening to detect viruses. This sequence represents a UL15
CC DNA packaging protein derived from VIDA relating to the novel method of
CC the invention
XX
SQ Sequence 351 AA;

Query Match 92.0%; Score 23; DB 6; Length 351;
Best Local Similarity 66.7%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db         53 DSLISC 58
```

RESULT 9
ABJ19477
ID ABJ19477 standard; protein; 351 AA.
XX
AC ABJ19477;
XX
DT 27-MAR-2003 (first entry)
XX
DE UL15 DNA packaging protein #24.
XX
KW Panel; degenerate primer pair; screening; virus family;
KW high throughput screening; UL15 DNA packaging protein; VIDA.
XX
OS Unidentified.
XX
PN WO200299129-A2.
XX
PD 12-DEC-2002.
XX
PF 07-JUN-2002; 2002WO-GB002640.
XX
PR 07-JUN-2001; 2001GB-00013908.
XX
PA (UNLO) UNIV COLLEGE LONDON.
XX
PI Griffiths DJ, Kellam P, Weiss RA;
XX
DR WPI; 2003-148676/14.
XX
PT Designing a panel of degenerate primer pairs, useful for screening new
PT members of multiple known virus families in a biological sample,
PT comprises deducing the sequences of the primers using computer based
PT calculations.
XX
PS Disclosure; Page 17; 31pp; English.
XX
CC The invention relates to a novel method for designing a panel of
CC degenerate primer pairs for screening for new members of multiple known
CC virus families in a biological sample. The method can be used in high
CC throughput screening to detect viruses. This sequence represents a UL15
CC DNA packaging protein derived from VIDA relating to the novel method of
CC the invention
XX
SQ Sequence 351 AA;

Query Match 92.0%; Score 23; DB 6; Length 351;
Best Local Similarity 66.7%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db         53 DSLISC 58
```

RESULT 10
ABU41410
ID ABU41410 standard; protein; 409 AA.
XX
AC ABU41410;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #26937.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas syringae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX

DR WPI; 2003-029926/02.
DR N-PSDB; ACA45280.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 69334; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 409 AA;

Query Match 92.0%; Score 23; DB 6; Length 409;
Best Local Similarity 66.7%; Pred. No. 2.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 233 DSLIAC 238

RESULT 11
ABM68320
ID ABM68320 standard; protein; 616 AA.
XX
AC ABM68320;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #1417.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 1417; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 616 AA;

Query Match 92.0%; Score 23; DB 6; Length 616;
Best Local Similarity 66.7%; Pred. No. 3e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 369 DTLIAC 374

RESULT 12
ADA55635
ID ADA55635 standard; protein; 628 AA.
XX
AC ADA55635;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 3203.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR N-PSDB; ADA53996.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 3203; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 628 AA;

Query Match 92.0%; Score 23; DB 6; Length 628;
Best Local Similarity 66.7%; Pred. No. 3.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| | | |
Db 326 DALITC 331

RESULT 13
ABP98882
ID ABP98882 standard; protein; 665 AA.
XX
AC ABP98882;
XX
DT 24-JUL-2003 (first entry)
XX
DE Human molecule for disease detection and treatment MDDT-10.
XX
KW Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; nephrotropic;
KW antithyroid; cerebroprotective; antiparkinsonian; anticonvulsant; MDDT;
KW nootropic; neuroprotective; antidiabetic; gene therapy; atherosclerosis;
KW molecule for disease detection and treatment; cancer; AIDS; allergy;
KW diabetes; glomerulonephritis; autoimmune thyroiditis; Cushing's syndrome;
KW stroke; Parkinson's disease; epilepsy.
XX
OS Homo sapiens.
XX
PN WO2003031595-A2.
XX
PD 17-APR-2003.
XX
PF 10-OCT-2002; 2002WO-US032852.
XX
PR 12-OCT-2001; 2001US-0328944P.
PR 26-OCT-2001; 2001US-0345384P.
PR 02-NOV-2001; 2001US-0343880P.
PR 09-NOV-2001; 2001US-0345143P.
PR 16-NOV-2001; 2001US-0332430P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Forsythe IJ, Emerling BM, Hafalia AJA, Yue H, Xu Y;
PI Gietzen KJ, Chawla NK, Baughn MR, Marquis JP, Becha SD, Kable AE;
PI Lal PG, Richardson TW, Lee SY, Lee EA, Tran B, Warren BA, Lu DAM;
PI Gururajan R, Sprague WW, Blake JJ, Thangavelu K, Swarnakar A;
PI Gorvad AE, Griffin JA, Lindquist EA, Elliott VS, Ison CH;
PI Ramkumar J;
XX
DR WPI; 2003-421277/39.
DR N-PSDB; ACC44397.
XX
PT Isolated peptide molecules for disease detection and treatment, useful
PT for diagnosing, treating or preventing disorders, e.g. cancer, AIDS,
PT atherosclerosis, diabetes or stroke.
XX
PS Claim 1; Page 156-157; 234pp; English.
XX
CC The invention relates to the isolation of a number of "molecules for
CC disease detection and treatment" (MDDT) and genes encoding them. The
CC invention also includes molecule which are at least 90% identical to the
CC protein and nucleotide sequences. This sequence represents a protein of
CC the invention. Disorders associated with aberrant expression of MDDT, are
CC cell proliferative disorders (e.g. cancer or atherosclerosis),
CC autoimmune/inflammatory disorders (e.g. AIDS, allergies, diabetes,
CC glomerulonephritis or autoimmune thyroiditis), developmental disorders
CC (e.g. Cushing's syndrome) or neurological disorders (e.g. stroke,
CC Parkinson's disease or epilepsy)
XX
SQ Sequence 665 AA;

Query Match 92.0%; Score 23; DB 6; Length 665;
Best Local Similarity 66.7%; Pred. No. 3.3e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| | | |
Db 363 DALITC 368

RESULT 14
ABJ19478
ID ABJ19478 standard; protein; 682 AA.
XX
AC ABJ19478;
XX
DT 27-MAR-2003 (first entry)
XX
DE UL15 DNA packaging protein #25.
XX
KW Panel; degenerate primer pair; screening; virus family;
KW high throughput screening; UL15 DNA packaging protein; VIDA.
XX
OS Unidentified.
XX
PN WO200299129-A2.
XX
PD 12-DEC-2002.
XX
PF 07-JUN-2002; 2002WO-GB002640.
XX
PR 07-JUN-2001; 2001GB-00013908.
XX
PA (UNLO) UNIV COLLEGE LONDON.
XX
PI Griffiths DJ, Kellam P, Weiss RA;
XX
DR WPI; 2003-148676/14.
XX
PT Designing a panel of degenerate primer pairs, useful for screening new
PT members of multiple known virus families in a biological sample,
PT comprises deducing the sequences of the primers using computer based
PT calculations.
XX
PS Disclosure; Page 17; 31pp; English.
XX
CC The invention relates to a novel method for designing a panel of
CC degenerate primer pairs for screening for new members of multiple known
CC virus families in a biological sample. The method can be used in high
CC throughput screening to detect viruses. This sequence represents a UL15

```
CC   DNA packaging protein derived from VIDA relating to the novel method of
CC   the invention
XX
SQ   Sequence 682 AA;

  Query Match             92.0%;  Score 23;  DB 6;  Length 682;
  Best Local Similarity   66.7%;  Pred. No. 3.3e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        385 DSLISC 390


RESULT 15
ABJ19521
ID   ABJ19521 standard; protein; 683 AA.
XX
AC   ABJ19521;
XX
DT   27-MAR-2003  (first entry)
XX
DE   UL15 DNA packaging protein #23.
XX
KW   Panel; degenerate primer pair; screening; virus family;
KW   high throughput screening; UL15 DNA packaging protein; VIDA.
XX
OS   Unidentified.
XX
PN   WO200299130-A2.
XX
PD   12-DEC-2002.
XX
PF   07-JUN-2002; 2002WO-GB002642.
XX
PR   07-JUN-2001; 2001GB-00013907.
XX
PA   (UNLO ) UNIV COLLEGE LONDON.
XX
PI   Griffiths DJ,  Kellam P,  Weiss RA;
XX
DR   WPI;:2003-148677/14.
XX
PT   High throughput method using degenerate polymerase chain reaction primers
PT   useful for screening a biological sample for unknown viruses.
XX
PS   Disclosure; Page 23-24; 39pp; English.
XX
CC   The invention relates to a novel method for designing a panel of
CC   degenerate primer pairs for screening for new members of multiple known
CC   virus families in a biological sample. The method can be used in high
CC   throughput screening to detect viruses. This sequence represents a UL15
CC   DNA packaging protein derived from VIDA relating to the novel method of
CC   the invention
XX
SQ   Sequence 683 AA;

  Query Match             92.0%;  Score 23;  DB 6;  Length 683;
  Best Local Similarity   66.7%;  Pred. No. 3.4e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        385 DSLISC 390


Search completed: March 31, 2004, 16:45:33
Job time : 40 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:42:17 ; Search time 10.6 Seconds
(without alignments)
29.222 Million cell updates/sec

Title: US-09-909-077-2
Perfect score: 25
Sequence: 1 DXLIXC 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22	88.0	18	4	US-08-469-260A-385	Sequence 385, App
2	22	88.0	18	4	US-08-488-446-385	Sequence 385, App
3	22	88.0	18	4	US-08-467-344A-385	Sequence 385, App
4	22	88.0	120	4	US-09-489-039A-12741	Sequence 12741, A
5	22	88.0	159	4	US-09-621-976-4449	Sequence 4449, Ap
6	22	88.0	163	4	US-09-489-039A-9040	Sequence 9040, Ap
7	22	88.0	237	4	US-09-489-039A-9365	Sequence 9365, Ap
8	22	88.0	264	3	US-08-969-644-16	Sequence 16, Appl
9	22	88.0	264	3	US-08-444-189-16	Sequence 16, Appl
10	22	88.0	264	3	US-08-468-544-16	Sequence 16, Appl
11	22	88.0	280	4	US-08-858-207A-267	Sequence 267, App
12	22	88.0	297	4	US-09-134-000C-4187	Sequence 4187, Ap
13	22	88.0	306	4	US-09-252-991A-30319	Sequence 30319, A
14	22	88.0	380	1	US-08-420-235B-5	Sequence 5, Appli
15	22	88.0	380	3	US-08-793-624-5	Sequence 5, Appli
16	22	88.0	380	4	US-09-134-001C-3583	Sequence 3583, Ap
17	22	88.0	380	5	PCT-US95-10194-5	Sequence 5, Appli
18	22	88.0	392	4	US-09-134-000C-6239	Sequence 6239, Ap
19	22	88.0	401	4	US-09-134-000C-4410	Sequence 4410, Ap
20	22	88.0	403	4	US-09-134-001C-5236	Sequence 5236, Ap
21	22	88.0	408	4	US-09-540-236-2740	Sequence 2740, Ap
22	22	88.0	413	4	US-09-252-991A-21766	Sequence 21766, A
23	22	88.0	419	4	US-09-543-681A-7295	Sequence 7295, Ap
24	22	88.0	423	4	US-09-328-352-5224	Sequence 5224, Ap
25	22	88.0	436	2	US-08-846-762-3	Sequence 3, Appli
26	22	88.0	436	2	US-08-846-762-72	Sequence 72, Appl
27	22	88.0	452	4	US-09-252-991A-31360	Sequence 31360, A
28	22	88.0	454	4	US-09-252-991A-28653	Sequence 28653, A
29	22	88.0	491	4	US-09-489-039A-8193	Sequence 8193, Ap
30	22	88.0	597	2	US-08-883-534-6	Sequence 6, Appli
31	22	88.0	597	3	US-09-204-764-6	Sequence 6, Appli
32	22	88.0	709	4	US-09-489-039A-11018	Sequence 11018, A
33	22	88.0	726	4	US-09-489-039A-7465	Sequence 7465, Ap
34	22	88.0	727	4	US-09-543-681A-7968	Sequence 7968, Ap
35	22	88.0	774	4	US-09-328-352-5361	Sequence 5361, Ap
36	22	88.0	872	1	US-08-491-357-3	Sequence 3, Appli
37	22	88.0	872	3	US-08-968-633-3	Sequence 3, Appli
38	22	88.0	872	3	US-09-196-466-3	Sequence 3, Appli
39	22	88.0	872	4	US-09-669-459A-3	Sequence 3, Appli
40	22	88.0	872	5	PCT-US96-10823-3	Sequence 3, Appli
41	22	88.0	1091	3	US-09-306-595C-7	Sequence 7, Appli
42	22	88.0	1091	4	US-09-925-388-7	Sequence 7, Appli
43	22	88.0	1207	4	US-09-976-594-4	Sequence 4, Appli
44	22	88.0	1493	4	US-09-489-039A-13687	Sequence 13687, A
45	21	84.0	8	3	US-09-258-754-142	Sequence 142, App

ALIGNMENTS

```
RESULT 1
US-08-469-260A-385
; Sequence 385, Application US/08469260A
; Patent No. 6451578
;  GENERAL INFORMATION:
;    APPLICANT:  JOHN N. SIMONS
;    APPLICANT:  TAMI J. PILOT-MATIAS
;    APPLICANT:  GEORGE J. DAWSON
;    APPLICANT:  GEORGE G. SCHLAUDER
;    APPLICANT:  SURESH M. DESAI
;    APPLICANT:  THOMAS P. LEARY
;    APPLICANT:  ANTHONY SCOTT MUERHOFF
;    APPLICANT:  JAMES C. ERKER
;    APPLICANT:  SHERI L. BUIJK
;    APPLICANT:  ISA K. MUSHAHWAR
;    TITLE OF INVENTION:  NON-A, NON-B. NON-C, NON-D, NON-E HEPATITIS
;    TITLE OF INVENTION:  REAGENTS AND METHODS FOR THEIR USE
;    NUMBER OF SEQUENCES:  716
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;      STREET:  100 ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,260A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/424,550
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5527.PC.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  385:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-08-469-260A-385

Query Match 88.0%; Score 22; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db          8 DQLITC 13
```

RESULT 2
US-08-488-446-385
; Sequence 385, Application US/08488446
; Patent No. 6558898
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSHAHWAR
; TITLE OF INVENTION: NON-A, NON-B. NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 385:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-446-385

Query Match 88.0%; Score 22; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db          8 DQLITC 13
```

RESULT 3

US-08-467-344A-385
; Sequence 385, Application US/08467344A
; Patent No. 6586568
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; TAMI J. PILOT-MATIAS
; GEORGE J. DAWSON
; GEORGE G. SCHLAUDER
; SURESH M. DESAI
; THOMAS P. LEARY
; ANTHONY SCOTT MUERHOFF
; JAMES C. ERKER
; SHERI L. BUIJK
; ISA K. MUSHAHWAR
; TITLE OF INVENTION: NON-A, NON-B. NON-C, NON-D, NON-E HEPATITIS
; REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,344A
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/424,550
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 385:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 385:
US-08-467-344A-385

Query Match 88.0%; Score 22; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db          8 DQLITC 13
```

RESULT 4
US-09-489-039A-12741
; Sequence 12741, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12741
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12741

Query Match 88.0%; Score 22; DB 4; Length 120;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 91 DDLIAC 96

RESULT 5
US-09-621-976-4449
; Sequence 4449, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4449
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4449

Query Match 88.0%; Score 22; DB 4; Length 159;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 88 DSLIQC 93

RESULT 6
US-09-489-039A-9040
; Sequence 9040, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9040
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9040

Query Match 88.0%; Score 22; DB 4; Length 163;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 78 DILIAC 83

RESULT 7
US-09-489-039A-9365
; Sequence 9365, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9365
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9365

Query Match 88.0%; Score 22; DB 4; Length 237;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 191 DLLIAC 196

RESULT 8
US-08-969-644-16
; Sequence 16, Application US/08969644
; Patent No. 6096519
; GENERAL INFORMATION:
; APPLICANT: Ratti, Giulio
; APPLICANT: Comanducci, Maurizio
; APPLICANT: Tecce, Mario F.
; APPLICANT: Giuliani, Marzia M.
; TITLE OF INVENTION: pCTD PLASMID ISOLATED FROM CHLAMYDIA
; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY
; TITLE OF INVENTION: THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID
; TITLE OF INVENTION: GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 301 N. Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22046-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/969,644
; FILING DATE: 13-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/467,152
; FILING DATE:
; APPLICATION NUMBER: US/07/661,820
; FILING DATE:
; APPLICATION NUMBER: IT MI 91A000314
; FILING DATE: 07-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 1267-202P
; TELECOMMUNICATION INFORMATION:

```
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;      TELEX:  248345
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  264 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-969-644-16

  Query Match             88.0%;  Score 22;  DB 3;  Length 264;
  Best Local Similarity   66.7%;  Pred. No. 4.8e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        142 DKLIAC 147


RESULT 9
US-08-444-189-16
; Sequence 16, Application US/08444189
; Patent No. 6110705
;  GENERAL INFORMATION:
;    APPLICANT:  Ratti, Giulio
;    APPLICANT:  Comanducci, Maurizio
;    APPLICANT:  Tecce, Mario F.
;    APPLICANT:  Giuliani, Marzia M.
;    TITLE OF INVENTION:  pCTD PLASMID ISOLATED FROM CHLAMYDIA
;    TITLE OF INVENTION:  TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY
;    TITLE OF INVENTION:  THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID
;    TITLE OF INVENTION:  GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  301 N. Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22046-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/444,189
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/180,528
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/991,512
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/661,820
;      FILING DATE:
;      APPLICATION NUMBER:  IT MI 91A000314
;      FILING DATE:  07-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svensson, Leonard R.
;      REGISTRATION NUMBER:  30,330
;      REFERENCE/DOCKET NUMBER:  1267-202P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  264 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-444-189-16

  Query Match             88.0%;  Score 22;  DB 3;  Length 264;
  Best Local Similarity   66.7%;  Pred. No. 4.8e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        142 DKLIAC 147


RESULT 10
US-08-468-544-16
; Sequence 16, Application US/08468544
; Patent No. 6248563
;  GENERAL INFORMATION:
;    APPLICANT:  Ratti, Giulio
;    APPLICANT:  Comanducci, Maurizio
;    APPLICANT:  Tecce, Mario F.
;    APPLICANT:  Giuliani, Marzia M.
;    TITLE OF INVENTION:  pCTD PLASMID ISOLATED FROM CHLAMYDIA
;    TITLE OF INVENTION:  TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY
;    TITLE OF INVENTION:  THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID
;    TITLE OF INVENTION:  GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  301 N. Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22046-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,544
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/661,820
;      FILING DATE:  28-FEB-1991
;      APPLICATION NUMBER:  IT MI 91A000314
;      FILING DATE:  07-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svensson, Leonard R.
;      REGISTRATION NUMBER:  30,330
;      REFERENCE/DOCKET NUMBER:  1267-202P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  264 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-468-544-16

  Query Match             88.0%;  Score 22;  DB 3;  Length 264;
  Best Local Similarity   66.7%;  Pred. No. 4.8e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        142 DKLIAC 147


RESULT 11
US-08-858-207A-267
```

```
; Sequence 267, Application US/08858207A
; Patent No. 6348328
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Stodola, Robert
;    TITLE OF INVENTION:  No. 6348328el Compounds
;    NUMBER OF SEQUENCES:  552
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,207A
;      FILING DATE:  09-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017670
;      FILING DATE:  14-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50475
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  267:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6348328e
US-08-858-207A-267

  Query Match             88.0%;  Score 22;  DB 4;  Length 280;
  Best Local Similarity   66.7%;  Pred. No. 5.1e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        213 DTLIIC 218


RESULT 12
US-09-134-000C-4187
; Sequence 4187, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4187
;   LENGTH: 297
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-4187

  Query Match             88.0%;  Score 22;  DB 4;  Length 297;
  Best Local Similarity   66.7%;  Pred. No. 5.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        143 DILITC 148


RESULT 13
US-09-252-991A-30319
; Sequence 30319, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 30319
;   LENGTH: 306
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30319

  Query Match             88.0%;  Score 22;  DB 4;  Length 306;
  Best Local Similarity   66.7%;  Pred. No. 5.5e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        166 DSLIFC 171


RESULT 14
US-08-420-235B-5
; Sequence 5, Application US/08420235B
; Patent No. 5801042
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/420,235B
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-B
;
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-420-235B-5

  Query Match             88.0%;  Score 22;  DB 1;  Length 380;
  Best Local Similarity   50.0%;  Pred. No. 6.8e+02;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | |: |
Db         82 DALVSC 87


RESULT 15
US-08-793-624-5
; Sequence 5, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
;  APPLICANT: Chang, Yuan
;  APPLICANT: Moore, Patrick S.
;  TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
;  TITLE OF INVENTION: Uses Thereof
;  FILE REFERENCE: 45185-C-PCT-US/JPW
;  CURRENT APPLICATION NUMBER: US/08/793,624C
;  CURRENT FILING DATE: 1997-02-18
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-5

  Query Match             88.0%;  Score 22;  DB 3;  Length 380;
  Best Local Similarity   50.0%;  Pred. No. 6.8e+02;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | |: |
Db         82 DALVSC 87


Search completed: March 31, 2004, 16:50:33
Job time : 11.6 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:45:43 ; Search time 25.8 Seconds
(without alignments)
60.852 Million cell updates/sec

Title: US-09-909-077-2
Perfect score: 25
Sequence: 1 DXLIXC 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	92.0	62	12	US-10-424-599-182620	Sequence 182620,
2	23	92.0	120	9	US-09-867-550-634	Sequence 634, App
3	23	92.0	263	12	US-10-424-599-229897	Sequence 229897,
4	23	92.0	351	9	US-09-780-053-5	Sequence 5, Appli
5	23	92.0	390	15	US-10-369-493-10089	Sequence 10089, A
6	23	92.0	409	12	US-10-282-122A-69334	Sequence 69334, A
7	23	92.0	628	15	US-10-094-749-3203	Sequence 3203, Ap
8	23	92.0	688	14	US-10-195-144-79	Sequence 79, Appl
9	23	92.0	688	15	US-10-345-072-79	Sequence 79, Appl
10	23	92.0	734	15	US-10-108-260A-3300	Sequence 3300, Ap
11	23	92.0	1207	10	US-09-949-029-144	Sequence 144, App
12	22	88.0	18	8	US-08-424-550B-385	Sequence 385, App
13	22	88.0	25	14	US-10-281-478-70	Sequence 70, Appl
14	22	88.0	29	9	US-09-864-761-38476	Sequence 38476, A
15	22	88.0	49	12	US-10-424-599-155408	Sequence 155408,
16	22	88.0	56	12	US-10-424-599-172462	Sequence 172462,
17	22	88.0	56	12	US-10-424-599-203805	Sequence 203805,
18	22	88.0	57	12	US-10-424-599-165494	Sequence 165494,
19	22	88.0	64	12	US-10-425-114-46652	Sequence 46652, A
20	22	88.0	72	12	US-10-424-599-175961	Sequence 175961,
21	22	88.0	80	12	US-10-424-599-180460	Sequence 180460,
22	22	88.0	84	12	US-10-424-599-167815	Sequence 167815,
23	22	88.0	86	12	US-10-424-599-220425	Sequence 220425,
24	22	88.0	99	12	US-10-425-114-50232	Sequence 50232, A
25	22	88.0	100	15	US-10-108-260A-2621	Sequence 2621, Ap
26	22	88.0	105	12	US-10-424-599-194421	Sequence 194421,
27	22	88.0	107	9	US-09-894-018-93	Sequence 93, Appl
28	22	88.0	107	12	US-10-424-599-273965	Sequence 273965,
29	22	88.0	108	12	US-10-424-599-230627	Sequence 230627,
30	22	88.0	117	14	US-10-029-386-31336	Sequence 31336, A
31	22	88.0	118	12	US-10-424-599-285104	Sequence 285104,
32	22	88.0	128	12	US-10-424-599-231269	Sequence 231269,
33	22	88.0	135	12	US-10-425-114-61321	Sequence 61321, A
34	22	88.0	135	12	US-10-425-114-61886	Sequence 61886, A
35	22	88.0	143	12	US-10-424-599-255242	Sequence 255242,
36	22	88.0	146	12	US-10-425-114-60556	Sequence 60556, A
37	22	88.0	152	12	US-10-424-599-266558	Sequence 266558,
38	22	88.0	164	14	US-10-029-386-33993	Sequence 33993, A
39	22	88.0	167	12	US-10-363-616-261	Sequence 261, App
40	22	88.0	184	12	US-10-424-599-186678	Sequence 186678,
41	22	88.0	187	11	US-09-801-944B-261	Sequence 261, App
42	22	88.0	187	12	US-10-424-599-218692	Sequence 218692,
43	22	88.0	197	11	US-09-864-408A-3394	Sequence 3394, Ap
44	22	88.0	204	11	US-09-801-944B-168	Sequence 168, App
45	22	88.0	208	12	US-10-282-122A-44819	Sequence 44819, A

ALIGNMENTS

RESULT 1
US-10-424-599-182620
; Sequence 182620, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 182620
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_135918C.1.pep
US-10-424-599-182620

Query Match 92.0%; Score 23; DB 12; Length 62;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db         16 DTLITC 21
```

RESULT 2
US-09-867-550-634
; Sequence 634, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.

```
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 634
;   LENGTH: 120
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-867-550-634

  Query Match             92.0%;  Score 23;  DB 9;  Length 120;
  Best Local Similarity   66.7%;  Pred. No. 4.8e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         17 DALIAC 22


RESULT 3
US-10-424-599-229897
; Sequence 229897, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 229897
;   LENGTH: 263
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_49620C.1.pep
US-10-424-599-229897

  Query Match             92.0%;  Score 23;  DB 12;  Length 263;
  Best Local Similarity   66.7%;  Pred. No. 1e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        256 DSLIAC 261


RESULT 4
US-09-780-053-5
; Sequence 5, Application US/09780053
; Patent No. US20020102640A1
; GENERAL INFORMATION:
;  APPLICANT: Rene S. Hubert
;  APPLICANT:  Daniel E.H. Afar
;  APPLICANT:  Pia M. Challita-Eid
;  APPLICANT:  Mary Faris
;  APPLICANT:  Elana Levin
;  APPLICANT:  Steve Chappell Mitchell
;  APPLICANT:  Aya Jakobovits
;  TITLE OF INVENTION: 83P5G4:  A TISSUE SPECIFIC PROTEIN
;  TITLE OF INVENTION:  HIGHLY EXPRESSED IN PROSTATE CANCER
;  FILE REFERENCE: 129.5USU1
;  CURRENT APPLICATION NUMBER: US/09/780,053
;  CURRENT FILING DATE:  2001-02-09
;  PRIOR APPLICATION NUMBER: 60/181,261
;  PRIOR FILING DATE: 2000-02-09
;  NUMBER OF SEQ ID NOS: 716
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 351
;   TYPE: PRT
;   ORGANISM: Drosophila Melanogaster
US-09-780-053-5

  Query Match             92.0%;  Score 23;  DB 9;  Length 351;
  Best Local Similarity   66.7%;  Pred. No. 1.3e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        185 DTLISC 190


RESULT 5
US-10-369-493-10089
; Sequence 10089, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10089
;   LENGTH: 390
;   TYPE: PRT
;   ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-10089

  Query Match             92.0%;  Score 23;  DB 15;  Length 390;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         27 DALIAC 32


RESULT 6
US-10-282-122A-69334
; Sequence 69334, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69334
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69334

Query Match 92.0%; Score 23; DB 12; Length 409;
Best Local Similarity 66.7%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 233 DSLIAC 238

RESULT 7
US-10-094-749-3203
; Sequence 3203, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3203
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-3203

Query Match 92.0%; Score 23; DB 15; Length 628;
Best Local Similarity 66.7%; Pred. No. 2.3e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 326 DALITC 331

RESULT 8
US-10-195-144-79
; Sequence 79, Application US/10195144
; Publication No. US20030126646A1
; GENERAL INFORMATION:
; APPLICANT: BROWN, GREGORY G.
; APPLICANT: FORMANOVA, NATASA
; APPLICANT: DENDY, CHARLES
; APPLICANT: LANDRY, BENOIT S.
; APPLICANT: CHEUNG, WING
; APPLICANT: JIN, HUA
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 16313-0136
; CURRENT APPLICATION NUMBER: US/10/195,144
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/305,026
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/305,363
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/308,736
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Raphanus sativum
US-10-195-144-79

Query Match 92.0%; Score 23; DB 14; Length 688;
Best Local Similarity 66.7%; Pred. No. 2.5e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 206 DTLISC 211

RESULT 9
US-10-345-072-79
; Sequence 79, Application US/10345072
; Publication No. US20030237112A1
; GENERAL INFORMATION:
; APPLICANT: BROWN, GREGORY G.
; APPLICANT: FORMANOVA, NATASA
; APPLICANT: DENDY, CHARLES
; APPLICANT: LANDRY, BENOIT S.
; APPLICANT: CHEUNG, WING
; APPLICANT: JIN, HUA
; APPLICANT: LAI, FANG MING
; APPLICANT: LEFOREST, MARTIN
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 16313-0210
; CURRENT APPLICATION NUMBER: US/10/345,072

```
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: PCT/US02/22217
;  PRIOR FILING DATE: 2002-07-12
;  PRIOR APPLICATION NUMBER: 60/305,026
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: 60/305,363
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: 60/308,736
;  PRIOR FILING DATE: 2001-07-30
;  NUMBER OF SEQ ID NOS: 179
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
;   LENGTH: 688
;   TYPE: PRT
;   ORGANISM: Raphanus sativum
US-10-345-072-79

  Query Match             92.0%;  Score 23;  DB 15;  Length 688;
  Best Local Similarity   66.7%;  Pred. No. 2.5e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        206 DTLISC 211


RESULT 10
US-10-108-260A-3300
; Sequence 3300, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3300
;   LENGTH: 734
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-3300

  Query Match             92.0%;  Score 23;  DB 15;  Length 734;
  Best Local Similarity   66.7%;  Pred. No. 2.6e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        129 DSLIAC 134


RESULT 11
US-09-949-029-144
; Sequence 144, Application US/09949029
; Publication No. US20030134278A1
; GENERAL INFORMATION:
;  APPLICANT: Karpen, G.H.
;  APPLICANT:  Dobie, K.W.
;  APPLICANT:  Kennedy, C.D.
;  APPLICANT:  Velasco, V.M.
;  APPLICANT:  McGrath, T.L.
;  APPLICANT:  Weko, J.
;  APPLICANT:  Patterson, R.W.
;  TITLE OF INVENTION: Identification of chromosome inheritance modifiers in Drosophila
;  TITLE OF INVENTION:  melanogaster
;  FILE REFERENCE: 1211.015US1
;  CURRENT APPLICATION NUMBER: US/09/949,029
;  CURRENT FILING DATE:  2001-09-07
;  PRIOR APPLICATION NUMBER: US 60/231,178
;  PRIOR FILING DATE: 2000-09-07
;  NUMBER OF SEQ ID NOS: 149
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 144
;   LENGTH: 1207
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-949-029-144

  Query Match             92.0%;  Score 23;  DB 10;  Length 1207;
  Best Local Similarity   66.7%;  Pred. No. 4.2e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        665 DALISC 670


RESULT 12
US-08-424-550B-385
; Sequence 385, Application US/08424550B
; Publication No. US20020119447A1
;  GENERAL INFORMATION:
;    APPLICANT:  JOHN N. SIMONS
;    APPLICANT:  TAMI J. PILOT-MATIAS
;    APPLICANT:  GEORGE J. DAWSON
;    APPLICANT:  GEORGE G. SCHLAUDER
;    APPLICANT:  SURESH M. DESAI
;    APPLICANT:  THOMAS P. LEARY
;    APPLICANT:  ANTHONY SCOTT MUERHOFF
;    APPLICANT:  JAMES C. ERKER
;    APPLICANT:  SHERI L. BUIJK
;    APPLICANT:  ISA K. MUSHAHWAR
;    TITLE OF INVENTION:  NON-A, NON-B. NON-C, NON-D, NON-E HEPATITIS
;    TITLE OF INVENTION:  REAGENTS AND METHODS FOR THEIR USE
;    NUMBER OF SEQUENCES:  716
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;      STREET:  100 ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/424,550B
;      FILING DATE:
;      CLASSIFICATION:  435435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5527.PC.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  385:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-424-550B-385

  Query Match             88.0%;  Score 22;  DB 8;  Length 18;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db          8 DQLITC 13
```

```
RESULT 13
US-10-281-478-70
; Sequence 70, Application US/10281478
; Publication No. US20030108959A1
; GENERAL INFORMATION:
;  APPLICANT: Immunex Corporation
;  APPLICANT:  Johnson, Richard S.
;  APPLICANT:  Guo, Lin
;  APPLICANT:  Mahimkar, Rajeev M.
;  APPLICANT:  Peschon, Jacques J.
;  APPLICANT:  Black, Roy A.
;  TITLE OF INVENTION: TREATING DISEASES MEDIATED BY METALLOPROTEASE-SHED PROTEINS
;  FILE REFERENCE: 3327-A
;  CURRENT APPLICATION NUMBER: US/10/281,478
;  CURRENT FILING DATE:  2002-10-25
;  NUMBER OF SEQ ID NOS: 158
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
;   LENGTH: 25
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: methionine sulfoxide
US-10-281-478-70

  Query Match             88.0%;  Score 22;  DB 14;  Length 25;
  Best Local Similarity   66.7%;  Pred. No. 1.9e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         15 DQLISC 20


RESULT 14
US-09-864-761-38476
; Sequence 38476, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38476
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC004149.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  26
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  22
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  25
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  24
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  23
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  23
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  24
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  21
US-09-864-761-38476

  Query Match             88.0%;  Score 22;  DB 9;  Length 29;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         20 DELIAC 25


RESULT 15
US-10-424-599-155408
; Sequence 155408, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 155408
;   LENGTH: 49
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_111354C.1.pep
US-10-424-599-155408

  Query Match             88.0%;  Score 22;  DB 12;  Length 49;
  Best Local Similarity   66.7%;  Pred. No. 3.6e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         42 DFLISC 47
```

Search completed: March 31, 2004, 16:52:56
Job time : 26.8 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:41:17 ; Search time 8.4 Seconds
(without alignments)
68.708 Million cell updates/sec

Title: US-09-909-077-2
Perfect score: 25
Sequence: 1 DXLIXC 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	92.0	100	2	D72577	hypothetical prote
2	23	92.0	213	2	B25750	nodulin-26b - soyb
3	23	92.0	262	2	C88325	protein F43G6.5 [i
4	23	92.0	310	2	T32283	hypothetical prote
5	23	92.0	356	2	B70424	lipid A disacchari
6	23	92.0	365	2	T50183	mpt1 like-protein
7	23	92.0	387	2	AG2939	conserved hypothet
8	23	92.0	387	2	H98342	hypothetical prote
9	23	92.0	414	1	C70859	probable hexosyltr
10	23	92.0	438	2	E87123	probable transfera
11	23	92.0	523	2	B55194	importin 2 - Afric
12	23	92.0	554	2	T22140	hypothetical prote
13	23	92.0	569	2	T00851	hypothetical prote
14	23	92.0	582	2	T39931	probable transcrip
15	23	92.0	603	2	T33134	hypothetical prote
16	23	92.0	656	2	B70766	hypothetical prote
17	23	92.0	683	1	QQBEM8	gene 29 protein -
18	23	92.0	683	2	T42943	hypothetical prote
19	23	92.0	758	2	S51748	lethal(2)denticlel
20	23	92.0	1118	2	S75309	hypothetical prote
21	23	92.0	1270	2	T30339	dsRNA adenosine de
22	23	92.0	1560	2	T30282	calcium-binding pr
23	23	92.0	1590	2	B87754	protein C43E11.3 [
24	22	88.0	50	2	S77838	hypothetical prote
25	22	88.0	80	2	A84528	hypothetical prote
26	22	88.0	98	2	A99856	hypothetical prote
27	22	88.0	113	1	S32873	hypA protein - Rhi
28	22	88.0	126	2	A38154	oogenesis required
29	22	88.0	143	2	G86698	transcription regu
30	22	88.0	148	2	AI1479	PTS system, enzyme
31	22	88.0	153	2	D95243	conserved hypothet
32	22	88.0	164	1	A40591	hydrogenase matura
33	22	88.0	173	2	S67579	probable membrane
34	22	88.0	174	2	G72381	hypothetical prote
35	22	88.0	179	2	B96989	probable membrane
36	22	88.0	180	2	E86134	hypothetical prote
37	22	88.0	180	2	B91293	hypothetical prote
38	22	88.0	190	2	T22279	hypothetical prote
39	22	88.0	200	2	C82224	probable alpha-rib
40	22	88.0	217	2	D87697	transaldolase fami
41	22	88.0	217	2	H87482	leucyl/phenylalany
42	22	88.0	220	2	S60210	fomA protein - Str
43	22	88.0	228	2	C97086	FNR-like catabolit
44	22	88.0	238	2	G97054	uncharacterized co
45	22	88.0	243	2	E37386	hypothetical prote

ALIGNMENTS

RESULT 1
D72577
hypothetical protein APE1900 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: D72577
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: D72577
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-100 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80905.1; PID:d1044691; PID:g51
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1900
C;Superfamily: Aeropyrum pernix hypothetical protein APE1900

Query Match 92.0%; Score 23; DB 2; Length 100;
Best Local Similarity 66.7%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              |||| |
Db         41 DTLITC 46
```

RESULT 2
B25750
nodulin-26b - soybean
C;Species: Glycine max (soybean)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 21-Jul-2000
C;Accession: B25750
R;Jacobs, F.A.; Zhang, M.; Fortin, M.G.; Verma, D.P.S.
Nucleic Acids Res. 15, 1271-1280, 1987
A;Title: Several nodulins of soybean share structural domains but differ in their subce
A;Reference number: A93653; MUID:87146431; PMID:3822823
A;Accession: B25750
A;Molecule type: mRNA
A;Residues: 1-213 <JAC>
A;Cross-references: GB:X05092; NID:g18707; PIDN:CAA28743.1; PID:g18708
A;Experimental source: cv. Prize
C;Superfamily: soybean nodulin-27

Query Match 92.0%; Score 23; DB 2; Length 213;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
```

```
              | || |
Db        120 DSLISC 125


RESULT 3
C88325
protein F43G6.5 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Sep-2003
C;Accession: C88325
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: C88325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <STO>
A;Cross-references: GB:chr_II; PIDN:CAA90398.1; PID:g3877085; GSPDB:GN00020; CESP:F43G6.
C;Genetics:
A;Gene: F43G6.5
A;Map position: 2
C;Superfamily: poly(A) polymerase

  Query Match             92.0%;  Score 23;  DB 2;  Length 262;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         54 DALITC 59


RESULT 4
T32283
hypothetical protein F40H7.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32283
R;Pauley, A.; Le, T.T.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid F40H7.
A;Reference number: Z21145
A;Accession: T32283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-310 <PAU>
A;Cross-references: EMBL:AF024498; PIDN:AAB70346.1; GSPDB:GN00020; CESP:F40H7.2
A;Experimental source: strain Bristol N2; clone F40H7
C;Genetics:
A;Gene: CESP:F40H7.2
A;Map position: 2
A;Introns: 17/3; 69/2; 94/1; 146/1; 203/3; 271/2

  Query Match             92.0%;  Score 23;  DB 2;  Length 310;
  Best Local Similarity   66.7%;  Pred. No. 1.9e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        171 DTLITC 176


RESULT 5
B70424
lipid A disaccharide synthetase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: B70424
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70424
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-356 <AQF>
A;Cross-references: GB:AE000740; NID:g2983826; PIDN:AAC07386.1; PID:g2983831; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: lpxB
C;Superfamily: lipid A disaccharide synthase

  Query Match             92.0%;  Score 23;  DB 2;  Length 356;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         82 DTLIAC 87


RESULT 6
T50183
mpt1 like-protein [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50183
R;Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 2000
A;Reference number: Z25044
A;Accession: T50183
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-365 <BAD>
A;Cross-references: EMBL:AL138854; PIDN:CAB72234.1; GSPDB:GN00066; SPDE:SPAC23G3.09
A;Experimental source: strain 972h(-); cosmid c23G3
C;Genetics:
A;Gene: SPDB:SPAC23G3.09
A;Map position: 1

  Query Match             92.0%;  Score 23;  DB 2;  Length 365;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         83 DALISC 88


RESULT 7
AG2939
conserved hypothetical protein Atu3117 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AG2939
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AG2939
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-387 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL43933.1; PID:g17741485; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3117
A;Map position: linear chromosome
```

Query Match 92.0%; Score 23; DB 2; Length 387;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db        112 DALIAC 117
```

RESULT 8
H98342
hypothetical protein AGR_L_3385 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98342
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98342
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-387 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90266.1; PID:g15160287; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3385
A;Map position: linear chromosome

Query Match 92.0%; Score 23; DB 2; Length 387;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db        112 DALIAC 117
```

RESULT 9
C70859
probable hexosyltransferase (EC 2.4.1.-) Rv3032 [similarity] - Mycobacterium tuberculosi
C;Species: Mycobacterium tuberculosis
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 16-Jun-2000
C;Accession: C70859
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: C70859
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-414 <COL>
A;Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16117.1; PID:g279163
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3032
C;Superfamily: probable hexosyltransferase ytxN
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 92.0%; Score 23; DB 1; Length 414;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db        169 DSLITC 174
```

RESULT 10
E87123
probable transferase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 17-May-2002
C;Accession: E87123
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: E87123
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-438 <STO>
A;Cross-references: GB:AL450380; NID:g13093472; PIDN:CAC30668.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML1715
C;Superfamily: probable hexosyltransferase ytxN

Query Match 92.0%; Score 23; DB 2; Length 438;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db        193 DSLITC 198
```

RESULT 11
B55194
importin 2 - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 07-May-1999
C;Accession: B55194
R;Goerlich, D.; Prehn, S.; Laskey, R.A.; Hartmann, E.
Cell 79, 767-778, 1994
A;Title: Isolation of a protein that is essential for the first step cf nuclear protein
A;Reference number: A55194; MUID:95094249; PMID:8001116
A;Accession: B55194
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-523 <GOE>
A;Cross-references: GB:L36340
A;Note: authors translated the codon AAT for residue 498 as Asp
C;Superfamily: pendulin

Query Match 92.0%; Score 23; DB 2; Length 523;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db        195 DALISC 200
```

RESULT 12
T22140
hypothetical protein F43G6.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Aug-2003
C;Accession: T22140
R;Swinburne, J.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19522
A;Accession: T22140
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-554 <WIL>
A;Cross-references: EMBL:Z50070; PIDN:CAA90398.2; GSPDB:GN00020; CESP:F43G6.5
A;Experimental source: clone F43G6
C;Genetics:

A;Gene: CESP:F43G6.5
A;Map position: 2
A;Introns: 191/3
C;Superfamily: poly(A) polymerase

Query Match 92.0%; Score 23; DB 2; Length 554;
Best Local Similarity 66.7%; Pred. No. 3.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 346 DALITC 351

RESULT 13
T00851
hypothetical protein At2g02770 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein T20F6.9
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C;Accession: T00851; F84440
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, March 1998
A;Description: Arabidopsis thaliana chromosome II BAC T20F6 genomic sequence.
A;Reference number: Z14206
A;Accession: T00851
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-569 <ROU>
A;Cross-references: EMBL:AC002521; NID:g2947056; PID:g2947064
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84440
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-569 <STO>
A;Cross-references: GB:AE002093; NID:g2947064; PIDN:AAC05345.1; GSPDB:GN00139
C;Genetics:
A;Gene: T20F6.9; At2g02770
A;Map position: 2
A;Introns: 34/1; 56/3; 172/1; 211/1; 295/3; 355/2; 414/1; 436/3; 470/3; 522/3

Query Match 92.0%; Score 23; DB 2; Length 569;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 458 DSLISC 463

RESULT 14
T39931
probable transcription factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39931
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Xiang, Z.; Aves, S.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21857
A;Accession: T39931
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-582 <LYN>
A;Cross-references: EMBL:AL023286; PIDN:CAA18865.1; GSPDB:GN00067; SPDB:SPBC21H7.05
A;Experimental source: strain 972h-; cosmid c21H7
C;Genetics:
A;Gene: SPDB:SPBC21H7.05
A;Map position: 2
A;Introns: 315/1; 354/1; 422/1

Query Match 92.0%; Score 23; DB 2; Length 582;
Best Local Similarity 66.7%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 328 DSLISC 333

RESULT 15
T33134
hypothetical protein C45G7.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 23-Sep-2002
C;Accession: T33134
R;Dante, M.; Wamsley, P.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C45G7.
A;Reference number: Z21288
A;Accession: T33134
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-603 <DAN>
A;Cross-references: EMBL:AF067611; PIDN:AAC19182.1; GSPDB:GN00022; CESP:C45G7.4
A;Experimental source: strain Bristol N2; clone C45G7
C;Genetics:
A;Gene: CESP:C45G7.4
A;Map position: 4
A;Introns: 60/3; 92/1; 200/3; 519/3
C;Superfamily: Caenorhabditis elegans protein C45G7.4; RING finger homology
F;20-78/Domain: RING finger homology <RRN>

Query Match 92.0%; Score 23; DB 2; Length 603;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DXLIXC 6
| ||| |
Db 19 DSLITC 24

Search completed: March 31, 2004, 16:49:30
Job time : 10.4 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:40 ; Search time 5.6 Seconds
(without alignments)
55.789 Million cell updates/sec

Title: US-09-909-077-2
Perfect score: 25
Sequence: 1 DXLIXC 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	92.0	219	1	N26B_SOYBN	P08863 glycine max
2	23	92.0	356	1	LPXB_AQUAE	O67420 aquifex aeo
3	23	92.0	397	1	TRPB_NITEU	Q82wi2 nitrosomona
4	23	92.0	409	1	TRPB_PSESH	Q849p2 pseudomonas
5	23	92.0	409	1	TRPB_PSESM	Q88b61 pseudomonas
6	23	92.0	522	1	IMA2_XENLA	P52171 xenopus lae
7	23	92.0	656	1	YK79_MYCTU	Q10687 mycobacteri
8	23	92.0	683	1	VTER_HSVSA	Q01020 herpesvirus
9	23	92.0	758	1	L2DT_DROME	Q24371 drosophila
10	23	92.0	2144	1	BP28_HUMAN	Q9h583 homo sapien
11	22	88.0	113	1	HYPA_RHILV	P28154 rhizobium l
12	22	88.0	126	1	PROF_DROME	P25843 drosophila
13	22	88.0	156	1	AURB_PIG	Q9n0x0 sus scrofa
14	22	88.0	217	1	LFTR_CAUCR	Q9a741 caulobacter
15	22	88.0	217	1	TAL_CAUCR	Q9a2f1 caulobacter
16	22	88.0	244	1	SRPB_YEAST	P36057 saccharomyc
17	22	88.0	249	1	GLO2_BUCBP	Q89an4 buchnera ap
18	22	88.0	264	1	GP5D_CHLTR	P10559 chlamydia t
19	22	88.0	286	1	HTPX_MYCTU	O06429 mycobacteri
20	22	88.0	287	1	HTPX_MYCLE	Q9cba4 mycobacteri
21	22	88.0	306	1	BUB2_YEAST	P26448 saccharomyc
22	22	88.0	308	1	CG17_YEAST	P25693 saccharomyc
23	22	88.0	316	1	FMT_CHLTR	O84535 chlamydia t
24	22	88.0	369	1	DHAS_BUCBP	Q89ab8 buchnera ap
25	22	88.0	379	1	TGT_STAAM	Q99t14 staphylococ
26	22	88.0	379	1	TGT_STAEP	Q8cml7 staphylococ
27	22	88.0	387	1	GALI_ENTFA	Q836p0 enterococcu
28	22	88.0	390	1	V084_FOWPV	O72902 fowlpox vir
29	22	88.0	391	1	TRB2_CHLCV	Q822w3 chlamydophi
30	22	88.0	392	1	TRPB_CHLTR	O84172 chlamydia t
31	22	88.0	394	1	TRPB_BACTN	Q8aad2 bacteroides
32	22	88.0	402	1	TRPB_LACLA	Q01998 lactococcus
33	22	88.0	402	1	TRPB_PSEAE	P07345 pseudomonas
34	22	88.0	402	1	TRPB_STAEP	Q8cpb1 staphylococ
35	22	88.0	403	1	TRPB_ACICA	P16706 acinetobact
36	22	88.0	405	1	TRPB_BRAJA	Q89we5 bradyrhizob
37	22	88.0	405	1	TRPB_PSEPK	Q88rp6 pseudomonas
38	22	88.0	405	1	TRPB_PSEPU	P11080 pseudomonas
39	22	88.0	407	1	TRPB_STRPN	Q97p32 streptococc
40	22	88.0	407	1	TRPB_STRR6	Q8dnm8 streptococc
41	22	88.0	408	1	TRPB_PSESY	P34817 pseudomonas
42	22	88.0	412	1	TRB1_CHLCV	Q822w9 chlamydophi
43	22	88.0	414	1	TRPB_GLOVI	Q7ngx9 gloeobacter
44	22	88.0	416	1	PGK_GLOMO	O74233 glomus moss
45	22	88.0	416	1	PGK_TRIRE	P14228 trichoderma

ALIGNMENTS

RESULT 1
N26B_SOYBN
ID N26B_SOYBN STANDARD; PRT; 219 AA.
AC P08863;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nodulin 26B precursor (N-26B).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Prize;
RX MEDLINE=87146431; PubMed=3822823;
RA Jacobs F.A., Zhang M., Fortin M.G., Verma D.P.S.;
RT "Several nodulins of soybean share structural domains but differ in
RT their subcellular locations.";
RL Nucleic Acids Res. 15:1271-1280(1987).
CC -!- INDUCTION: During nodulation in legume roots after Rhizobium
CC infection.
CC -!- SIMILARITY: Belongs to the nodulin 20 family.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-4 is the initiator.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 5.
CC ---

CC ---
DR EMBL; X05092; CAA28743.1; ALT_FRAME.
DR InterPro; IPR003387; Nodulin.
DR Pfam; PF02451; Nodulin; 1.
KW Nodulation; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 219 NODULIN 26B.
SQ SEQUENCE 219 AA; 23786 MW; E3696853E6D44A75 CRC64;

Query Match 92.0%; Score 23; DB 1; Length 219;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | ||| |
Db        126 DSLISC 131
```

RESULT 2
LPXB_AQUAE
ID LPXB_AQUAE STANDARD; PRT; 356 AA.

```
AC   O67420;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Lipid-A-disaccharide synthase (EC 2.4.1.182).
GN   LPXB OR AQ_1427.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: Condensation of UDP-2,3-diacylglucosamine and 2,3-
CC       diacylglucosamine-1-phosphate to form lipid A disaccharide, a
CC       precursor of lipid A, a phosphorylated glycolipid that anchors the
CC       lipopolysaccharide to the outer membrane of the cell (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: UDP-2,3-bis(3-hydroxytetradecanoyl)glucosamine
CC       + 2,3-bis(3-hydroxytetradecanoyl)-beta-D-glucosaminyl 1-phosphate
CC       = UDP + 2,3-bis(3-hydroxytetradecanoyl)-D-glucosaminyl-1,6-beta-D-
CC       2,3-bis(3-hydroxytetradecanoyl)-beta-D-glucosaminyl 1-phosphate.
CC   -!- PATHWAY: Lipid A biosynthesis; fifth step.
CC   -!- SIMILARITY: Belongs to the lpxB family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000740; AAC07386.1; -.
DR   PIR; B70424; B70424.
DR   HAMAP; MF_00392; -; 1.
DR   InterPro; IPR003835; Glyco_trans_19.
DR   Pfam; PF02684; LpxB; 1.
DR   TIGRFAMs; TIGR00215; lpxB; 1.
KW   Transferase; Glycosyltransferase; Lipid A biosynthesis;
KW   Lipid synthesis; Complete proteome.
SQ   SEQUENCE   356 AA;  41300 MW;  1B4CFBA5F409CD68 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 356;
  Best Local Similarity   66.7%;  Pred. No. 75;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         82 DTLIAC 87


RESULT 3
TRPB_NITEU
ID   TRPB_NITEU     STANDARD;      PRT;   397 AA.
AC   Q82WI2;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Tryptophan synthase beta chain (EC 4.2.1.20).
GN   TRPB OR NE0693.
OS   Nitrosomonas europaea.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC   Nitrosomonadaceae; Nitrosomonas.
OX   NCBI_TaxID=915;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19718 / IFO 14298;
RX   MEDLINE=22586410; PubMed=12700255;
RA   Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA   Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA   Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT   "Complete genome sequence of the ammonia-oxidizing bacterium and
RT   obligate chemolithoautotroph Nitrosomonas europaea.";
RL   J. Bacteriol. 185:2759-2773(2003).
CC   -!- FUNCTION: The beta subunit is responsible for the synthesis of L-
CC       tryptophan from indole and L-serine.
CC   -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC       = L-tryptophan + glyceraldehyde 3-phosphate.
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC   -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the trpB family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BX321858; CAD84604.1; -.
DR   HAMAP; MF_00133; -; 1.
DR   InterPro; IPR001926; B6_enzyme_beta.
DR   InterPro; IPR006654; Trp_synth_beta.
DR   InterPro; IPR006653; Trp_synth_b_rel.
DR   Pfam; PF00291; PALP; 1.
DR   TIGRFAMs; TIGR00263; trpB; 1.
DR   PROSITE; PS00168; TRP_SYNTHASE_BETA; 1.
KW   Tryptophan biosynthesis; Pyridoxal phosphate; Lyase;
KW   Complete proteome.
FT   BINDING      90     90       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   397 AA;  43399 MW;  8BD08F3CA35827B4 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 397;
  Best Local Similarity   66.7%;  Pred. No. 84;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        228 DALIAC 233


RESULT 4
TRPB_PSESH
ID   TRPB_PSESH     STANDARD;      PRT;   409 AA.
AC   Q849P2;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Tryptophan synthase beta chain (EC 4.2.1.20).
GN   TRPB.
OS   Pseudomonas syringae (pv. phaseolicola).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=319;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Race 7 isolate 1449;
RA   Tsaltas D.;
RT   "Biochemical, structural and molecular characterization of resistant
RT   interactions between Pseudomonas syringae pv. phaseolicola and
RT   Phaseolus vulgaris.";
RL   Thesis (2003), University of London, U.K.
CC   -!- FUNCTION: The beta subunit is responsible for the synthesis of L-
CC       tryptophan from indole and L-serine.
CC   -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC       = L-tryptophan + glyceraldehyde 3-phosphate.
```

```
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC   -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the trpB family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY210847; AAO50076.1; -.
DR   HAMAP; MF_00133; -; 1.
DR   InterPro; IPR001926; B6_enzyme_beta.
DR   InterPro; IPR006654; Trp_synth_beta.
DR   InterPro; IPR006653; Trp_synth_b_rel.
DR   Pfam; PF00291; PALP; 1.
DR   TIGRFAMs; TIGR00263; trpB; 1.
DR   PROSITE; PS00168; TRP_SYNTHASE_BETA; 1.
KW   Tryptophan biosynthesis; Pyridoxal phosphate; Lyase.
FT   BINDING      95      95       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   409 AA;  44541 MW;  D6311A4BB6C9BABF CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 409;
  Best Local Similarity   66.7%;  Pred. No. 87;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        233 DSLIAC 238


RESULT 5
TRPB_PSESM
ID   TRPB_PSESM     STANDARD;      PRT;   409 AA.
AC   Q88B61;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Tryptophan synthase beta chain (EC 4.2.1.20).
GN   TRPB OR PSPTO0158.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=22834015; PubMed=12928499;
RA   Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA   Gwinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA   Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA   Nelson W.C., Davidsen T., Zafar N., Zhou L., Liu J., Yuan Q.,
RA   Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
RA   Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
RA   Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P.,
RA   Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
RA   White O., Fraser C.M., Collmer A.;
RT   "The complete genome sequence of the Arabidopsis and tomato pathogen
RT   Pseudomonas syringae pv. tomato DC3000.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC   -!- FUNCTION: The beta subunit is responsible for the synthesis of L-
CC       tryptophan from indole and L-serine.
CC   -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC       = L-tryptophan + glyceraldehyde 3-phosphate.
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC   -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the trpB family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016856; AAO53712.1; -.
DR   TIGR; PSPTO0158; -.
DR   HAMAP; MF_00133; -; 1.
DR   InterPro; IPR001926; B6_enzyme_beta.
DR   Pfam; PF00291; PALP; 1.
DR   PROSITE; PS00168; TRP_SYNTHASE_BETA; 1.
KW   Tryptophan biosynthesis; Pyridoxal phosphate; Lyase;
KW   Complete proteome.
FT   BINDING      95      95       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   409 AA;  44546 MW;  25A56962380CC284 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 409;
  Best Local Similarity   66.7%;  Pred. No. 87;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        233 DSLIAC 238


RESULT 6
IMA2_XENLA
ID   IMA2_XENLA     STANDARD;      PRT;   522 AA.
AC   P52171;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Importin alpha-2 subunit (Karyopherin alpha-2 subunit).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-54; 295-331; 341-345 AND
RP   506-522.
RC   TISSUE=Ovary;
RX   MEDLINE=95094249; PubMed=8001116;
RA   Goerlich D., Prehn S., Laskey R.A., Hartmann E.;
RT   "Isolation of a protein that is essential for the first step of
RT   nuclear protein import.";
RL   Cell 79:767-778(1994).
CC   -!- FUNCTION: IT IS ESSENTIAL FOR SELECTIVE PROTEIN IMPORT INTO
CC       NUCLEUS. PROMOTES SIGNAL-DEPENDENT BINDING OF KARYOPHILIC PROTEINS
CC       TO THE NUCLEAR ENVELOPE.
CC   -!- SUBUNIT: Forms a complex with importin beta-1 subunit.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the importin alpha family.
CC   -!- SIMILARITY: Contains 8 ARM repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L36340; AAC14196.1; -.
DR   HSSP; Q02821; 1BK5.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR000225; Armadillo.
DR   InterPro; IPR002652; ImportinA_B.
DR   Pfam; PF00514; Armadillo_seg; 8.
```

```
DR   Pfam; PF01749; IBB; 1.
DR   SMART; SM00185; ARM; 8.
DR   PROSITE; PS50176; ARM_REPEAT; 3.
KW   Transport; Protein transport; Repeat.
FT   INIT_MET      0      0
FT   DOMAIN        9     49       IBB.
FT   REPEAT      108    150       ARM 1.
FT   REPEAT      151    195       ARM 2.
FT   REPEAT      196    234       ARM 3.
FT   REPEAT      235    279       ARM 4.
FT   REPEAT      280    319       ARM 5.
FT   REPEAT      320    361       ARM 6.
FT   REPEAT      362    401       ARM 7.
FT   REPEAT      402    446       ARM 8.
FT   DOMAIN      447    522       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT    308    308       T -> P (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   522 AA;  57753 MW;  97EFFD5292300410 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 522;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        194 DALISC 199


RESULT 7
YK79_MYCTU
ID   YK79_MYCTU     STANDARD;      PRT;   656 AA.
AC   Q10687;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein Rv2079/MT2140.
GN   RV2079 OR MT2140 OR MTCY49.18.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- SIMILARITY: SOME, TO M.TUBERCULOSIS RV0963C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98192.1; -.
DR   EMBL; AE007063; AAK46422.1; -.
DR   PIR; B70766; B70766.
DR   TIGR; MT2140; -.
DR   TuberculList; Rv2079; -.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT     47     47       Y -> C (IN REF. 2).
SQ   SEQUENCE   656 AA;  69823 MW;  9DFB74A58809D3E4 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 656;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        155 DALITC 160


RESULT 8
VTER_HSVSA
ID   VTER_HSVSA     STANDARD;      PRT;   683 AA.
AC   Q01020;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Probable DNA packaging protein.
GN   29.
OS   Herpesvirus saimiri (strain 11).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=10383;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92333688; PubMed=1321287;
RA   Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA   Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA   Honess R.W.;
RT   "Primary structure of the herpesvirus saimiri genome.";
RL   J. Virol. 66:5047-5058(1992).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL15,
CC       HSV-6 ORF12L, EHV-1 44, HCMV UL89, EBV BGRF1/BDRF1, AND VZV 42/45.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64346; CAA45657.1; -.
DR   InterPro; IPR003498; DNA_pack_C.
DR   InterPro; IPR003499; DNA_pack_N.
DR   Pfam; PF02499; DNA_pack_C; 1.
DR   Pfam; PF02500; DNA_pack_N; 1.
KW   DNA packaging.
SQ   SEQUENCE   683 AA;  77049 MW;  1B66CC277156AFC4 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 683;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        385 DSLISC 390


RESULT 9
L2DT_DROME
```

```
ID   L2DT_DROME     STANDARD;      PRT;   758 AA.
AC   Q24371;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Lethal(2)denticleless protein (DTL83 protein).
GN   L(2)DTL OR DTL83.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=96257214; PubMed=8666267;
RA   Kurzik-Dumke U., Neubauer M., Debes A.;
RT   "Identification of a novel Drosophila melanogaster heat-shock gene,
RT   lethal(2)denticleless [l(2)dtl], coding for an 83-kDa protein.";
RL   Gene 171:163-170(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-2;
RA   Gunacker S., Neubhuer M., Kurzik-Dumke U.;
RT   "Sequence of a 10291 nt genomic region of Drosophila melanogaster
RT   harbouring the genes l(2)tid, l(2)not, l(2)rot, and l(2)dtl.";
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED DURING EMBRYOGENESIS
CC       WITH NO SIGN OF TISSUE SPECIFICITY IN EXPRESSION UP TO STAGE 17.
CC   -!- DEVELOPMENTAL STAGE: DETECTED AT ALL DEVELOPMENTAL STAGES. THE
CC       EXTREMELY HIGH LEVEL OF TRANSCRIPTION DETECTED IN THE EARLY EMBRYO
CC       AND IN ADULTS IS CAUSED BY MATERNAL MESSAGE.
CC   -!- INDUCTION: By heat shock.
CC   -!- SIMILARITY: Contains 6 WD repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83414; CAA58441.1; -.
DR   EMBL; X98094; CAA66723.1; -.
DR   PIR; S51748; S51748.
DR   FlyBase; FBgn0013548; l(2)dtl.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 5.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00320; WD40; 5.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 5.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Heat shock; Repeat; WD repeat.
FT   REPEAT       88    118       WD 1.
FT   REPEAT      132    163       WD 2.
FT   REPEAT      183    238       WD 3.
FT   REPEAT      253    292       WD 4.
FT   REPEAT      309    338       WD 5.
FT   REPEAT      351    382       WD 6.
FT   DOMAIN      558    567       POLY-ALA.
FT   DOMAIN      672    675       POLY-GLY.
FT   DOMAIN      721    724       POLY-THR.
SQ   SEQUENCE   758 AA;  82352 MW;  3A05FEF78D0502F1 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 758;
  Best Local Similarity   66.7%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        222 DTLISC 227


RESULT 10
BP28_HUMAN
ID   BP28_HUMAN     STANDARD;      PRT;  2144 AA.
AC   Q9H583; Q9NW23;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein BAP28.
GN   BAP28.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
RP   GLY-2017.
RA   Bougueleret L., Chumakov I., Barry C., Cohen-Akenine A.;
RT   "A novel BAP28 gene and protein.";
RL   Patent number WO0100669, 04-JAN-2001.
RN   [2]
RP   SEQUENCE OF 1534-2144 FROM N.A.
RA   Cobley V.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1777-2144 FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the BAP28 family.
CC   -!- SIMILARITY: Contains 1 HEAT repeat.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AX067150; CAC26776.1; -.
DR   EMBL; AL136105; CAC15948.1; -.
DR   EMBL; AK001221; BAA91564.1; ALT_INIT.
DR   SWISS-2DPAGE; Q9H583; HUMAN.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR000357; HEAT.
DR   PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
KW   Polymorphism.
FT   REPEAT     2106   2142       HEAT.
FT   VARIANT    1694   1694       N -> S.
FT                                /FTId=VAR_010939.
FT   VARIANT    1854   1854       V -> A.
FT                                /FTId=VAR_010940.
FT   VARIANT    1967   1967       N -> D.
FT                                /FTId=VAR_010941.
FT   VARIANT    2017   2017       E -> G.
FT                                /FTId=VAR_010942.
SQ   SEQUENCE   2144 AA;  242355 MW;  D66816EE78D8C9B7 CRC64;

  Query Match             92.0%;  Score 23;  DB 1;  Length 2144;
  Best Local Similarity   66.7%;  Pred. No. 4.7e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
```

```
Db        129 DSLIAC 134


RESULT 11
HYPA_RHILV
ID   HYPA_RHILV     STANDARD;      PRT;   113 AA.
AC   P28154;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hydrogenase nickel incorporation protein hypA (Protein hupL).
GN   HYPA OR HUPL.
OS   Rhizobium leguminosarum (biovar viciae).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX   NCBI_TaxID=387;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=128c53;
RX   MEDLINE=93316844; PubMed=8326860;
RA   Rey L., Murillo J., Hernando Y., Hidalgo E., Cabrera E., Imperial J.,
RA   Ruiz-Argueeso T.;
RT   "Molecular analysis of a microaerobically induced operon required for
RT   hydrogenase synthesis in Rhizobium leguminosarum biovar viciae.";
RL   Mol. Microbiol. 8:471-481(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10;
RA   Brito B., Palacios J.M., Imperial J., Ruiz-Argueso T., Yang W.C.,
RA   Bisseling T., Schmitt H., Kerl V., Bauer T., Kokotek W., Lotz W.;
RT   "Organization of the hup-region and its differential transcription
RT   in non-symbiotic and symbiotic cells of Rhizobium leguminosarum
RT   bv. viciae B10.";
RL   Mol. Plant Microbe Interact. 8:235-240(1997).
CC   -!- FUNCTION: Probably plays a role in an hydrogenase nickel cofactor
CC       insertion step.
CC   -!- SIMILARITY: Belongs to the hypA/hybF family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52974; CAA37159.1; -.
DR   EMBL; Z36981; CAA85441.1; -.
DR   PIR; S32873; S32873.
DR   HAMAP; MF_00213; -; 1.
DR   InterPro; IPR000688; HypA.
DR   Pfam; PF01155; HypA; 1.
DR   ProDom; PD003620; HypA; 1.
DR   TIGRFAMs; TIGR00100; hypA; 1.
DR   PROSITE; PS01249; HYPA; 1.
KW   Metal-binding; Nickel.
FT   METAL         2      2       NICKEL (POTENTIAL).
FT   METAL        73     73       NICKEL (POTENTIAL).
FT   METAL        76     76       NICKEL (POTENTIAL).
FT   METAL        89     89       NICKEL (POTENTIAL).
FT   METAL        92     92       NICKEL (POTENTIAL).
SQ   SEQUENCE   113 AA;  12447 MW;  F657134559CDAC67 CRC64;

  Query Match             88.0%;  Score 22;  DB 1;  Length 113;
  Best Local Similarity   66.7%;  Pred. No. 44;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db         42 DALIFC 47


RESULT 12
```

```
PROF_DROME
ID   PROF_DROME     STANDARD;      PRT;   126 AA.
AC   P25843; Q9VMJ9;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Profilin (Chickadee protein).
GN   CHIC OR CHI OR CG9553.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92208942; PubMed=1339308;
RA   Cooley L., Verheyen E., Caverly K.;
RT   "Chickadee encodes a profilin required for intercellular cytoplasm
RT   transport during Drosophila oogenesis.";
RL   Cell 69:173-184(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: Binds to actin and affects the structure of the
CC       cytoskeleton. At high concentrations, profilin prevents the
CC       polymerization of actin, whereas it enhances it at low
CC       concentrations. By binding to PIP2, it inhibits the formation of
CC       IP3 and DG. This profilin is required for intercellular cytoplasm
CC       transport during Drosophila oogenesis.
CC   -!- SUBUNIT: Occurs in many kinds of cells as a complex with monomeric
CC       actin in a 1:1 ratio.
CC   -!- SIMILARITY: Belongs to the profilin family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M84528; AAA28418.1; -.
DR   EMBL; M84529; AAA28419.1; -.
DR   EMBL; AE003612; AAF52315.1; -.
DR   PIR; A38154; A38154.
DR   HSSP; P07763; 1ACF.
DR   FlyBase; FBgn0000308; chic.
DR   GO; GO:0007300; P:nurse cell/oocyte transport (sensu Insecta); IMP.
DR   InterPro; IPR002097; Profilin.
DR   InterPro; IPR005455; Profilin_plant.
DR   Pfam; PF00235; profilin; 1.
DR   PRINTS; PR00392; PROFILIN.
DR   PRINTS; PR01640; PROFILINPLNT.
DR   SMART; SM00392; PROF; 1.
DR   PROSITE; PS00414; PROFILIN; 1.
KW   Actin-binding; Cytoskeleton.
SQ   SEQUENCE   126 AA;  13723 MW;  6E2942C12A81ADEA CRC64;

  Query Match             88.0%;  Score 22;  DB 1;  Length 126;
  Best Local Similarity   66.7%;  Pred. No. 49;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        119 DYLITC 124


RESULT 13
AURB_PIG
ID   AURB_PIG       STANDARD;      PRT;   156 AA.
AC   Q9N0X0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Serine/threonine protein kinase 12 (EC 2.7.1.37) (Aurora-B)
DE   (Fragment).
GN   AURKB OR STK12.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Small intestine;
RX   MEDLINE=22232238; PubMed=12270407;
RA   Braun F., Hosseini S.M., Lorf T., Laabs S., Ringe B.;
RT   "Differential gene expression during intestinal ischemia-reperfusion
RT   injury.";
RL   Transplant. Proc. 34:2301-2302(2002).
CC   -!- FUNCTION: May be directly involved in regulating the cleavage of
CC       polar spindle microtubules and is a key regulator for the onset of
CC       cytokinesis during mitosis (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBCELLULAR LOCATION: Localized to the midzone of central spindle
CC       in late anaphase and concentrated into the midbody in telophase
CC       and cytokinesis. Colocalized with gamma tubulin in the mid-body
CC       (By similarity).
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC       Aurora subfamily.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF244364; AAF61735.1; -.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Cell cycle; Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   NON_TER       1      1
FT   DOMAIN        <1    >156       PROTEIN KINASE.
FT   ACT_SITE      51     51        BY SIMILARITY.
FT   NON_TER      156    156
SQ   SEQUENCE   156 AA;  17867 MW;  CD23040EDB633FCE CRC64;

  Query Match             88.0%;  Score 22;  DB 1;  Length 156;
  Best Local Similarity   66.7%;  Pred. No. 61;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db         37 DALIYC 42


RESULT 14
LFTR_CAUCR
ID   LFTR_CAUCR     STANDARD;      PRT;   217 AA.
AC   Q9A741;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Leucyl/phenylalanyl-tRNA--protein transferase (EC 2.3.2.6) (L/F-
DE   transferase) (Leucyltransferase) (Phenyalanyltransferase).
GN   AAT OR CC1885.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Functions in the N-end rule pathway of protein
CC       degradation where it conjugates Leu, Phe and, less efficiently,
CC       Met from aminoacyl-tRNAs to the N-termini of proteins containing
CC       an N-terminal arginine or lysine (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-leucyl-tRNA + protein = tRNA + L-leucyl-
CC       protein.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanyl-tRNA + protein = tRNA + L-
CC       phenylalany-protein.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the L/F-transferase family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE005862; AAK23860.1; -.
```

```
DR   PIR; H87482; H87482.
DR   TIGR; CC1885; -.
DR   HAMAP; MF_00688; -; 1.
DR   InterPro; IPR004616; Aat.
DR   Pfam; PF03588; Leu_Phe_trans; 1.
DR   TIGRFAMs; TIGR00667; aat; 1.
KW   Transferase; Acyltransferase; Complete proteome.
SQ   SEQUENCE   217 AA;  23746 MW;  C16D2599926D840A CRC64;

  Query Match             88.0%;  Score 22;  DB 1;  Length 217;
  Best Local Similarity   66.7%;  Pred. No. 85;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db          8 DDLIAC 13


RESULT 15
TAL_CAUCR
ID   TAL_CAUCR      STANDARD;      PRT;   217 AA.
AC   Q9A2F1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Proabable transaldolase (EC 2.2.1.2).
GN   TAL OR CC3614.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Transaldolase is important for the balance of
CC       metabolites in the pentose-phosphate pathway (By similarity).
CC   -!- CATALYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glyceraldehyde
CC       3-phosphate = D-erythrose 4-phosphate + D-fructose 6-phosphate.
CC   -!- PATHWAY: Pentose phosphate pathway; nonoxidative part.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: Belongs to the transaldolase family. Subfamily 3B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006020; AAK25576.1; -.
DR   PIR; D87697; D87697.
DR   HSSP; P30148; 1ONR.
DR   TIGR; CC3614; -.
DR   HAMAP; MF_00494; -; 1.
DR   InterPro; IPR001585; Transaldolase.
DR   InterPro; IPR004731; Transaldolase_C.
DR   Pfam; PF00923; Transaldolase; 1.
DR   TIGRFAMs; TIGR00875; talC; 1.
DR   PROSITE; PS01054; TRANSALDOLASE_1; 1.
DR   PROSITE; PS00958; TRANSALDOLASE_2; FALSE_NEG.
KW   Transferase; Pentose shunt; Complete proteome.
FT   ACT_SITE     83     83       BY SIMILARITY.
SQ   SEQUENCE   217 AA;  23029 MW;  E600551B85DCE65B CRC64;

  Query Match             88.0%;  Score 22;  DB 1;  Length 217;
  Best Local Similarity   66.7%;  Pred. No. 85;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db         89 DGLIAC 94


Search completed: March 31, 2004, 16:46:14
Job time : 7.6 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:41:17 ; Search time 11.2 Seconds
(without alignments)
68.708 Million cell updates/sec

Title: US-09-909-077-1
Perfect score: 32
Sequence: 1 DTEDVVXX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	93.8	412	2	T15214	hypothetical prote
2	30	93.8	424	1	A36000	sperm-binding glyc
3	30	93.8	439	2	G88103	protein W10G11.17
4	30	93.8	2135	2	T14602	variant-specific s
5	30	93.8	3011	1	GNWVCH	genome polyprotein
6	29	90.6	371	1	QQVZA7	A7L protein - vacc
7	29	90.6	372	2	D72164	A7L protein - vari
8	29	90.6	372	2	H42517	A6L protein - vacc
9	29	90.6	372	2	G36848	A6L protein - vari
10	29	90.6	372	2	T28548	hypothetical prote
11	29	90.6	372	2	T37393	probable 43.1K pro
12	28	87.5	326	2	C90187	hypothetical prote
13	28	87.5	695	2	T52429	PRM1 homolog [impo
14	28	87.5	793	2	T31655	DNA excision repai
15	28	87.5	899	2	H96617	probable disease r
16	27	84.4	52	2	PN0481	cell division cycl
17	27	84.4	171	2	S77242	hypothetical prote
18	27	84.4	181	2	B87138	probable adenylate
19	27	84.4	181	2	T45390	adenylate kinase [
20	27	84.4	304	2	S21342	ring-infected eryt
21	27	84.4	333	2	E97257	spore coat protein
22	27	84.4	380	2	F89811	cystathionine gamm
23	27	84.4	382	2	H90127	Mrs2p [imported] -
24	27	84.4	426	2	S70396	zona pellucida gly
25	27	84.4	434	2	AI0337	tetrahydrofolylpol
26	27	84.4	435	2	S57668	hypothetical prote
27	27	84.4	463	2	D95172	peptidoglycan N-ac
28	27	84.4	463	2	D98038	peptidoglycan GlcN
29	27	84.4	590	2	S57594	hypothetical prote
30	27	84.4	720	2	T44562	catalase (EC 1.11.
31	27	84.4	725	2	E96596	hypothetical prote
32	27	84.4	824	2	AB2829	DNA helicase II [i
33	27	84.4	827	2	H97606	probable DNA helic
34	27	84.4	831	2	B40204	Na+/H+-exchanging
35	27	84.4	885	2	T26310	hypothetical prote
36	27	84.4	999	2	A34307	Ca2+-transporting
37	27	84.4	1383	2	C96723	hypothetical prote
38	27	84.4	1876	2	C70749	probable ppsA prot
39	26	81.2	74	2	B90467	conserved hypothet
40	26	81.2	82	2	T18159	hypothetical prote
41	26	81.2	111	2	C84100	cytochrome c551 pr
42	26	81.2	121	2	A85806	unknown protein en
43	26	81.2	121	2	H90957	hypothetical prote
44	26	81.2	124	2	B83822	cytochrome c551 cc
45	26	81.2	129	2	H71156	hypothetical prote

ALIGNMENTS

RESULT 1
T15214
hypothetical protein F57C9.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15214
R;Geisel, C.; Kramer, J.; Gibson, A.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid F57C9.
A;Reference number: Z18309
A;Accession: T15214
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-412 <GEI>
A;Cross-references: EMBL:AF003142; NID:g2088743; PID:g2088750; PIDN:AAB54190.1; GSPDB:GI
A;Experimental source: strain Bristol N2; clone F57C9
C;Genetics:
A;Gene: CESP:F57C9.7
A;Map position: 1
A;Introns: 100/3; 158/3; 194/3; 272/2; 311/2

Query Match 93.8%; Score 30; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||||
Db        132 DTEDVV 137
```

RESULT 2
A36000
sperm-binding glycoprotein ZP3 precursor - human
N;Alternate names: sperm receptor ZP3; zona pellucida glycoprotein ZP3
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36000; A44365
R;Chamberlin, M.E.; Dean, J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6014-6018, 1990
A;Title: Human homolog of the mouse sperm receptor.
A;Reference number: A36000; MUID:90349545; PMID:2385582
A;Accession: A36000
A;Molecule type: mRNA; DNA
A;Residues: 1-424 <CHA>
A;Cross-references: GB:M60504; GB:M35109; NID:g340491; PIDN:AAA61336.1; PID:g340492
R;van Duin, M.; Polman, J.E.; Verkoelen, C.C.; Bunschoten, H.; Meyerink, J.H.; Olijve, W
Genomics 14, 1064-1070, 1992
A;Title: Cloning and characterization of the human sperm receptor ligand ZP3: evidence f
A;Reference number: A44365; MUID:93122771; PMID:1478648
A;Accession: A44365
A;Status: preliminary
A;Molecule type: mRNA

A;Residues: 329-370,'S',372-424 <VAN>
A;Experimental source: ovary
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:122391, NCBIP:122392)
C;Comment: This sulfated glycoprotein in the zona pellucida of the oocyte is a receptor
C;Genetics:
A;Gene: GDB:ZP3A
A;Cross-references: GDB:128007; OMIM:182889
A;Map position: 7pter-7qter
C;Superfamily: sperm-binding glycoprotein ZP3; ZP domain homology
C;Keywords: glycoprotein; oocyte; receptor; sulfoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-424/Product: sperm-binding glycoprotein ZP3 #status predicted <MAT>
F;45-301/Domain: ZP domain homology <ZPH>

Query Match 93.8%; Score 30; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 85 DTEDVV 90

RESULT 3
G88103
protein W10G11.17 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: G88103
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: G88103
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-439 <STO>
A;Cross-references: GB:chr_II; PIDN:AB95081.1; PID:g2746939; GSPDB:GN00020; CESP:W10G11.
C;Genetics:
A;Gene: W10G11.17
A;Map position: 2

Query Match 93.8%; Score 30; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 40 DTEDVV 45

RESULT 4
T14602
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C;Accession: T14602
R;Voss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998
A;Description: Identification of a conserved 5' flanking region of Plasmodium falciparum
A;Reference number: Z18158
A;Accession: T14602
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2135 <VOS>
A;Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1
C;Genetics:
A;Gene: varph17

Query Match 93.8%; Score 30; DB 2; Length 2135;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1324 DTEDVV 1329

RESULT 5
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 93.8%; Score 30; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 2413 DTEDVV 2418

RESULT 6
QQVZA7
A7L protein - vaccinia virus (strain WR)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 08-Apr-1994
C;Accession: C41806
R;Ahn, B.Y.; Rosel, J.; Cole, N.B.; Moss, B.
J. Virol. 66, 971-982, 1992
A;Title: Identification and expression of rpo19, a vaccinia virus gene encoding a 19-kil
A;Reference number: A41806; MUID:92114202; PMID:1731116
A;Accession: C41806
A;Molecule type: DNA
A;Residues: 1-371 <AHN>
A;Cross-references: GB:M76473
C;Superfamily: vaccinia virus A7L protein

```
  Query Match             90.6%;  Score 29;  DB 1;  Length 371;
  Best Local Similarity   83.3%;  Pred. No. 73;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        126 DTEDIV 131


RESULT 7
D72164
A7L protein - variola minor virus (strain Garcia-1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: D72164
R;Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of alastrim variola minor
A;Reference number: A72150
A;Accession: D72164
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <SHC>
A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54710.1; PID:g5830671
A;Experimental source: strain Garcia-1966
C;Genetics:
A;Gene: A7L
C;Superfamily: vaccinia virus A7L protein

  Query Match             90.6%;  Score 29;  DB 2;  Length 372;
  Best Local Similarity   83.3%;  Pred. No. 73;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132


RESULT 8
H42517
A6L protein - vaccinia virus (strain Copenhagen)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 08-Apr-1994
C;Accession: H42517
R;Johnson, G.P.
submitted to GenBank, June 1990
A;Reference number: A33172
A;Accession: H42517
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <JOH>
C;Superfamily: vaccinia virus A7L protein

  Query Match             90.6%;  Score 29;  DB 2;  Length 372;
  Best Local Similarity   83.3%;  Pred. No. 73;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132


RESULT 9
G36848
A6L protein - variola virus (strain India-1967)
C;Species: variola virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C;Accession: G36848; S46892
R;Blinov, V.M.
submitted to GenBank, November 1992
A;Reference number: A36859
A;Accession: G36848
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CAA49051.1; PID:g297289
R;Volchov, V.E.; Blinov, V.M.; Totmenin, A.V.; Shchelkunov, S.N.; Sandakhchiev, L.S.
submitted to the EMBL Data Library, April 1992
A;Description: Nucleotide sequence analysis of the region of variola virus XhoI-G genom
A;Reference number: S46890
A;Accession: S46892
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <VOL>
A;Cross-references: EMBL:X67116; NID:g516451; PIDN:CAA47514.1; PID:g516454
C;Superfamily: vaccinia virus A7L protein

  Query Match             90.6%;  Score 29;  DB 2;  Length 372;
  Best Local Similarity   83.3%;  Pred. No. 73;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132


RESULT 10
T28548
hypothetical protein A7L - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28548
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubi
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of varicla smallpox virus
A;Reference number: Z20488; MUID:94088747; PMID:8264798
A;Accession: T28548
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-372 <MAS>
A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60858.1; PID:g439028
A;Experimental source: strain Bangladesh-1975
C;Superfamily: vaccinia virus A7L protein

  Query Match             90.6%;  Score 29;  DB 2;  Length 372;
  Best Local Similarity   83.3%;  Pred. No. 73;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132


RESULT 11
T37393
probable 43.1K protein - vaccinia virus (strain Ankara)
C;Species: vaccinia virus
A;Variety: strain Ankara
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C;Accession: T37393
R;Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.
submitted to the EMBL Data Library, March 1997
A;Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strai
A;Reference number: Z20877
A;Accession: T37393
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-372 <ANT>
A;Cross-references: EMBL:U94848; PIDN:AAB96459.1
A;Experimental source: strain Ankara
C;Genetics:
A;Note: MVA117L
C;Superfamily: vaccinia virus A7L protein

  Query Match             90.6%;  Score 29;  DB 2;  Length 372;
```

Best Local Similarity 83.3%; Pred. No. 73;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132
```

RESULT 12
C90187
hypothetical protein SSO0429 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: C90187
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: C90187
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-326 <KUR>
A;Cross-references: GB:AE006641; NID:g13813580; PIDN:AAK40754.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO0429

Query Match 87.5%; Score 28; DB 2; Length 326;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||::
Db        297 DTEDII 302
```

RESULT 13
T52429
PRM1 homolog [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C;Accession: T52429
R;Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.
Gene 239, 309-316, 1999
A;Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic reg
A;Reference number: Z25171
A;Accession: T52429
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-695 <KAT>
A;Cross-references: EMBL:AB028231; PIDN:BAA87956.1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: CW9
A;Map position: 1

Query Match 87.5%; Score 28; DB 2; Length 695;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||::
Db         71 DTEDII 76
```

RESULT 14
T31655
DNA excision repair cross-complementing protein - sea squirt (Ciona intestinalis)
C;Species: Ciona intestinalis
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Nov-2000
C;Accession: T31655
R;Bird, A.P.; Clark, V.; Jones, S.J.; Leitgeb, S.; Dobson, R.; Tweedie, S.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z21049
A;Accession: T31655
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-793 <BIR>
A;Cross-references: EMBL:Z83760; PIDN:CAB06045.1
C;Genetics:
A;Note: COS41.2
C;Superfamily: DNA excision repair cross-complementing protein ERCC3

Query Match 87.5%; Score 28; DB 2; Length 793;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||::
Db        138 DTEDII 143
```

RESULT 15
H96617
probable disease resistance protein F9K23.6 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96617
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96617
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-899 <STO>
A;Cross-references: GB:AE005173; NID:g11034973; PIDN:AAG27128.1; GSPDB:GN00141
C;Genetics:
A;Gene: F9K23.6
A;Map position: 1

Query Match 87.5%; Score 28; DB 2; Length 899;
Best Local Similarity 66.7%; Pred. No. 3.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||::
Db         69 DTEDII 74
```

Search completed: March 31, 2004, 16:49:28
Job time : 13.2 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:40 ; Search time 7.46667 Seconds
(without alignments)
55.789 Million cell updates/sec

Title: US-09-909-077-1
Perfect score: 32
Sequence: 1 DTEDVVXX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	93.8	424	1	ZP3_CALSQ	P53786 callithrix
2	30	93.8	424	1	ZP3_HUMAN	P21754 homo sapien
3	30	93.8	424	1	ZP3_MACRA	P53785 macaca radi
4	30	93.8	3011	1	POLG_HCVH	P27958 h genome po
5	29	90.6	371	1	VA06_VACCV	P29192 vaccinia vi
6	29	90.6	372	1	VA06_VACCC	P20985 vaccinia vi
7	29	90.6	372	1	VA06_VARV	P33833 variola vir
8	29	90.6	459	1	TRME_STAEP	Q8cmn5 staphylococ
9	28	87.5	899	1	R24L_ARATH	Q9c646 arabidopsis
10	27	84.4	181	1	KAD_MYCLE	O33007 mycobacteri
11	27	84.4	286	1	3HAO_RAT	P46953 rattus norv
12	27	84.4	304	1	RESA_PLAFP	Q26005 plasmodium
13	27	84.4	335	1	KG62_MOUSE	Q8c010 mus musculu
14	27	84.4	426	1	ZP3_CANFA	P48831 canis famil
15	27	84.4	590	1	YM72_YEAST	Q05021 saccharomyc
16	27	84.4	720	1	CATA_HALN1	O73955 halobacteri
17	27	84.4	760	1	RESA_PLAFN	P13831 plasmodium
18	27	84.4	764	1	SMZ2_BRARE	Q9w6g6 brachydanio
19	27	84.4	831	1	NAH3_RAT	P26433 rattus norv
20	27	84.4	953	1	B3A4_RAT	Q8k4v2 rattus norv
21	27	84.4	999	1	ATA3_RAT	P18596 rattus norv
22	27	84.4	1038	1	ATA3_MOUSE	Q64518 mus musculu
23	27	84.4	1876	1	PPSA_MYCTU	Q10977 mycobacteri
24	26	81.2	105	1	FLHD_RALSO	Q8xq73 ralstonia s
25	26	81.2	181	1	KAD_MYCTU	O53796 mycobacteri
26	26	81.2	184	1	RAPA_DISOM	P22123 discopyge o
27	26	81.2	184	1	RAPA_HUMAN	P10113 homo sapien
28	26	81.2	205	1	F18B_HUMAN	Q9nyz1 homo sapien
29	26	81.2	205	1	F18B_MOUSE	Q9d8t4 mus musculu
30	26	81.2	211	1	KAD_FUSNN	Q8re31 fusobacteri
31	26	81.2	234	1	NK4_HUMAN	P24001 homo sapien
32	26	81.2	237	1	GP44_BPSP1	Q08402 bacteriopha
33	26	81.2	253	1	SOJ_BACHD	Q9k5n0 bacillus ha
34	26	81.2	253	1	Y990_CAMJE	P45489 campylobact
35	26	81.2	284	1	KADC_ARATH	Q9zuu1 arabidopsis
36	26	81.2	297	1	PH4H_CHRVO	P30967 chromobacte
37	26	81.2	364	1	MYCL_HUMAN	P12524 homo sapien
38	26	81.2	368	1	MYCL_MOUSE	P10166 mus musculu
39	26	81.2	369	1	YN65_ARCFU	O30305 archaeoglob
40	26	81.2	371	1	LRC2_MOUSE	Q8vdb8 mus musculu
41	26	81.2	393	1	YBZJ_CAEEL	P91119 caenorhabdi
42	26	81.2	405	1	YIS1_STRCO	P19780 streptomyce
43	26	81.2	420	1	Z229_HUMAN	Q9ujw7 homo sapien
44	26	81.2	430	1	K117_HUMAN	P42696 homo sapien
45	26	81.2	458	1	TRME_CLOPE	Q8xh30 clostridium

ALIGNMENTS

RESULT 1
ZP3_CALSQ
ID ZP3_CALSQ STANDARD; PRT; 424 AA.
AC P53786;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor).
GN ZP3.
OS Callithrix sp. (Marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC Callithrix.
OX NCBI_TaxID=9485;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=94363314; PubMed=8081814;
RA Thillai-Koothan P., van Duin M., Aitken R.J.;
RT "Cloning, sequencing and oocyte-specific expression of the marmoset
RT sperm receptor protein, ZP3.";
RL Zygote 1:93-101(1993).
CC -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC sperm-adhesion to the zona pellucida, and may contribute to the
CC species-specificity of the insemination.
CC -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC matrix.
CC -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC -!- SIMILARITY: Contains 1 ZP domain.
CC ---

CC ---
DR EMBL; S71825; AAB31866.1; -.
DR InterPro; IPR001507; Endoglin/CD105.
DR Pfam; PF00100; zona_pellucida; 1.
DR PRINTS; PR00023; ZPELLUCIDA.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS00682; ZP_DOMAIN; 1.
KW Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW Extracellular matrix; Multigene family.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 424 ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT DOMAIN 23 387 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 388 408 POTENTIAL.
FT DOMAIN 409 424 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 45 307 ZP.
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  46809 MW;  1DACBD03026C2739 CRC64;

  Query Match             93.8%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 2
ZP3_HUMAN
ID   ZP3_HUMAN      STANDARD;      PRT;   424 AA.
AC   P21754; Q06633;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor)
DE   (ZP3A/ZP3B).
GN   ZP3 OR ZP3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ZP3A).
RX   MEDLINE=90349545; PubMed=2385582;
RA   Chamberlin M.E., Dean J.;
RT   "Human homolog of the mouse sperm receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6014-6018(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS ZP3A AND ZP3B).
RC   TISSUE=Ovary;
RX   MEDLINE=93122771; PubMed=1478648;
RA   van Duin M., Polman J.E., Verkoelen C.C., Bunschoten H.,
RA   Meyerink J.H., Olijve W., Aitken R.J.;
RT   "Cloning and characterization of the human sperm receptor ligand ZP3:
RT   evidence for a second polymorphic allele with a different frequency
RT   in the Caucasian and Japanese populations.";
RL   Genomics 14:1064-1070(1992).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN WHICH
CC       ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=ZP3A;
CC         IsoId=P21754-1; Sequence=Displayed;
CC       Name=ZP3B;
CC         IsoId=P21754-2; Sequence=VSP_006949, VSP_006950;
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60504; AAA61336.1; -.
DR   EMBL; X56777; CAA40095.1; -.
DR   EMBL; A18567; CAA01398.1; -.
DR   PIR; A36000; A36000.
DR   Genew; HGNC:13189; ZP3.
DR   MIM; 182889; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family; Alternative splicing.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    388    408       POTENTIAL.
FT   DOMAIN      409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    307       ZP.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    364    372       PLIFLDRRG -> ATDLPGQEW (in isoform ZP3B).
FT                                 /FTId=VSP_006949.
FT   VARSPLIC    373    424       Missing (in isoform ZP3B).
FT                                 /FTId=VSP_006950.
FT   CONFLICT    345    345       R -> T (IN REF. 2).
SQ   SEQUENCE   424 AA;  47028 MW;  94517B7B66EFFE06 CRC64;

  Query Match             93.8%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 3
ZP3_MACRA
ID   ZP3_MACRA      STANDARD;      PRT;   424 AA.
AC   P53785;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor).
GN   ZP3.
OS   Macaca radiata (Bonnet monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9548;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96249321; PubMed=8848588;
RA   Kolluri S.K., Kaul R., Banerjee K., Gupta S.K.;
RT   "Nucleotide sequence of cDNA encoding bonnet monkey (Macaca radiata)
RT   zona pellucida glycoprotein-ZP3.";
RL   Reprod. Fertil. Dev. 7:1209-1212(1995).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X82639; CAA57961.1; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    388    408       POTENTIAL.
FT   DOMAIN      409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    307       ZP.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  47040 MW;  3B41C4CFA3792331 CRC64;

  Query Match             93.8%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 4
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
```

```
FT   CHAIN        747    809       PROTEIN P7.
FT   CHAIN        810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN       1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN       1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN       1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN       1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN       2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM     347    369       POTENTIAL.
FT   ACT_SITE    1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND     1230   1237       ATP (POTENTIAL).
FT   SITE        1316   1319       DECH BOX.
FT   CARBOHYD     196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND      1224   1226
FT   TURN        1232   1233
FT   TURN        1236   1238
FT   HELIX       1239   1246
FT   TURN        1247   1248
FT   STRAND      1251   1255
FT   HELIX       1258   1271
FT   TURN        1272   1272
FT   STRAND      1277   1280
FT   TURN        1281   1282
FT   STRAND      1283   1285
FT   STRAND      1291   1295
FT   HELIX       1296   1301
FT   TURN        1302   1303
FT   STRAND      1312   1316
FT   TURN        1317   1319
FT   HELIX       1323   1335
FT   TURN        1336   1340
FT   STRAND      1343   1347
FT   TURN        1352   1353
FT   TURN        1360   1361
FT   STRAND      1362   1366
FT   STRAND      1368   1368
FT   STRAND      1373   1375
FT   TURN        1376   1377
FT   STRAND      1378   1380
FT   HELIX       1382   1385
FT   STRAND      1389   1393
FT   HELIX       1397   1409
FT   TURN        1410   1411
FT   STRAND      1414   1417
FT   TURN        1419   1420
FT   STRAND      1432   1436
FT   TURN        1438   1439
FT   STRAND      1450   1453
FT   STRAND      1456   1463
FT   STRAND      1471   1478
FT   STRAND      1480   1480
FT   HELIX       1481   1488
FT   TURN        1489   1490
FT   STRAND      1497   1501
FT   STRAND      1507   1507
FT   STRAND      1511   1511
FT   HELIX       1514   1527
FT   HELIX       1532   1544
FT   STRAND      1550   1550
FT   HELIX       1555   1564
FT   HELIX       1570   1578
FT   TURN        1579   1580
FT   HELIX       1584   1597
FT   TURN        1598   1598
FT   HELIX       1606   1611
FT   TURN        1614   1618
FT   STRAND      1622   1623
FT   STRAND      1627   1627
FT   STRAND      1635   1636
FT   HELIX       1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match             93.8%;  Score 30;  DB 1;  Length 3011;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db       2413 DTEDVV 2418


RESULT 5
VA06_VACCV
ID   VA06_VACCV     STANDARD;      PRT;   371 AA.
AC   P29132;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A6 (A7).
GN   A6L OR A7L.
OS   Vaccinia virus (strain WR).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114202; PubMed=1731116;
RA   Ahn B.-Y., Rosel J., Cole N.B., Moss B.;
RT   "Identification and expression of rpo19, a vaccinia virus gene
RT   encoding a 19-kilodalton DNA-dependent RNA polymerase subunit.";
RL   J. Virol. 66:971-982(1992).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76473; -; NOT_ANNOTATED_CDS.
DR   PIR; C41806; QQVZA7.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   371 AA;  43006 MW;  013A21BD94612AC6 CRC64;

  Query Match             90.6%;  Score 29;  DB 1;  Length 371;
  Best Local Similarity   83.3%;  Pred. No. 38;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        126 DTEDIV 131


RESULT 6
VA06_VACCC
ID   VA06_VACCC     STANDARD;      PRT;   372 AA.
AC   P20985;
DT   01-FEB-1991 (Rel. 17, Created)
```

```
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A6.
GN   A6L.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35027; AAA48123.1; -.
DR   PIR; H42517; H42517.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43127 MW;  8149CD07AD808D70 CRC64;

  Query Match             90.6%;  Score 29;  DB 1;  Length 372;
  Best Local Similarity   83.3%;  Pred. No. 38;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132


RESULT 7
VA06_VARV
ID   VA06_VARV      STANDARD;      PRT;   372 AA.
AC   P33833;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A6.
GN   A6L OR A7L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=India-1967 / Isolate Ind3;
RX   MEDLINE=92209372; PubMed=1666548;
RA   Shchelkunov S.N., Marennikova S.S., Totmenin A.V., Blinov V.M.,
RA   Chizhikov V.E., Gutorov V.V., Safronov P.F., Pozdnyakov S.G.,
RA   Shelukhina E.M., Gashnikov P.V., Anjaparidze O.G., Sandakhchiev L.S.;
RT   "Creation of a clone library of fragments from the natural variola
RT   virus and study of the structural and functional organization of
RT   viral genes from a circle of hosts.";
RL   Dokl. Akad. Nauk SSSR 321:402-406(1991).
RN   [2]
RP   COMPLETE GENOME.
RC   STRAIN=India-1967 / Isolate Ind3;
RX   MEDLINE=93202281; PubMed=8384129;
RA   Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT   "Genes of variola and vaccinia viruses necessary to overcome the host
RT   protective mechanisms.";
RL   FEBS Lett. 319:80-83(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bangladesh-1975;
RX   MEDLINE=94088747; PubMed=8264798;
RA   Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA   Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA   Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA   Venter C.J.;
RT   "Potential virulence determinants in terminal regions of variola
RT   smallpox virus genome.";
RL   Nature 366:748-751(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Garcia-1966;
RA   Shchelkunov S.N., Totmenin A.V., Safronov P.F., Resenchuk S.M.,
RA   Blinov V.M., Sandakhchiev L.S.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69198; CAA49051.1; -.
DR   EMBL; X67116; CAA47514.1; -.
DR   EMBL; L22579; AAA60858.1; -.
DR   EMBL; X76265; CAA53848.1; -.
DR   PIR; D72164; D72164.
DR   PIR; G36848; G36848.
DR   PIR; T28548; T28548.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43148 MW;  9751173CE363452D CRC64;

  Query Match             90.6%;  Score 29;  DB 1;  Length 372;
  Best Local Similarity   83.3%;  Pred. No. 38;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        127 DTEDIV 132


RESULT 8
TRME_STAEP
ID   TRME_STAEP     STANDARD;      PRT;   459 AA.
AC   Q8CMN5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   tRNA modification GTPase trmE.
GN   TRME OR SE2417.
OS   Staphylococcus epidermidis.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1282;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12228;
RX   PubMed=12950922;
RA   Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA   Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA   Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT   "Genome-based analysis of virulence genes in a non-biofilm-forming
RT   Staphylococcus epidermidis strain (ATCC 12228).";
```

```
RL   Mol. Microbiol. 49:1577-1593(2003).
CC   -!- FUNCTION: Exhibits a very high intrinsic GTPase hydrolysis rate.
CC       Involved in the biosynthesis of the hypermodified nucleoside 5-
CC       methylaminomethyl-2-thiouridine, which is found in the wobble
CC       position of some tRNAs (By similarity).
CC   -!- SIMILARITY: Belongs to the era/trmE family of GTP-binding
CC       proteins. TrmE subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016752; AAO06060.1; -.
DR   HAMAP; MF_00379; -; 1.
DR   InterPro; IPR005289; GTP-bindding_dom.
DR   InterPro; IPR002917; MMR_HSR1.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR004520; ThdF.
DR   Pfam; PF01926; MMR_HSR1; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   TIGRFAMs; TIGR00650; MG442; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   TIGRFAMs; TIGR00450; thdF; 1.
KW   tRNA processing; GTP-binding; Complete proteome.
FT   NP_BIND     228    235       GTP (POTENTIAL).
FT   NP_BIND     275    279       GTP (POTENTIAL).
FT   NP_BIND     335    338       GTP (POTENTIAL).
SQ   SEQUENCE   459 AA;  51451 MW;  BE425B02FE9D2AA1 CRC64;

  Query Match             90.6%;  Score 29;  DB 1;  Length 459;
  Best Local Similarity   83.3%;  Pred. No. 48;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        281 DTEDIV 286


RESULT 9
R24L_ARATH
ID   R24L_ARATH     STANDARD;      PRT;   899 AA.
AC   Q9C646; Q9SM04;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable disease resistance protein RXW24L.
GN   RXW24L OR AT1G58410 OR X7J.10 OR F9K23.6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [2]
RP   SEQUENCE OF 584-899 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20018182; PubMed=10548732;
RA   Kato A., Suzuki M., Kuwahara A., Ooe H., Higano-Inaba K., Komeda Y.;
RT   "Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana
RT   genomic region located around the 100 map unit of chromosome 1.";
RL   Gene 239:309-316(1999).
CC   -!- FUNCTION: Potential disease resistance protein.
CC   -!- DOMAIN: The LRR repeats probably act as specificity determinant of
CC       pathogen recognition (By similarity).
CC   -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC   -!- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC   -!- SIMILARITY: Contains 1 NB-ARC domain.
CC   -!- DATABASE: NAME=NIB-LRRS;
CC       NOTE=Functional and comparative genomics of disease resistance gene
CC       homologs;
CC       WWW="http://niblrrs.ucdavis.edu".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC082643; AAG50648.1; -.
DR   EMBL; AB077822; BAB83873.1; -.
DR   EMBL; AB008019; BAA88266.1; -.
DR   PIR; H96617; H96617.
DR   InterPro; IPR000767; Disease_resist.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR002182; NB-ARC.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF00931; NB-ARC; 1.
DR   PRINTS; PR00364; DISEASERSIST.
KW   Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
FT   DOMAIN       8     43       LEUCINE-ZIPPER.
FT   DOMAIN     143    455       NB-ARC.
FT   REPEAT     601    624       LRR 1.
FT   REPEAT     840    864       LRR 2.
FT   NP_BIND    189    196       ATP (POTENTIAL).
SQ   SEQUENCE   899 AA;  104332 MW;  99DC7FF6D1A5335F CRC64;

  Query Match             87.5%;  Score 28;  DB 1;  Length 899;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||::
Db         69 DTEDII 74


RESULT 10
KAD_MYCLE
ID   KAD_MYCLE      STANDARD;      PRT;   181 AA.
AC   O33007;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN   ADK OR ML1832 OR MLCB2492.28.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
```

```
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: This small ubiquitous enzyme is essential for
CC       maintenance and cell growth.
CC   -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the adenylate kinase family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z98756; CAB11460.1; -.
DR   EMBL; AL583923; CAC30786.1; -.
DR   PIR; B87138; B87138.
DR   PIR; T45390; T45390.
DR   HSSP; P27142; 1ZIN.
DR   Leproma; ML1832; -.
DR   HAMAP; MF_00235; -; 1.
DR   InterPro; IPR006259; Adenyl_kin_sub.
DR   InterPro; IPR000850; Adenylate_kin.
DR   Pfam; PF00406; ADK; 1.
DR   PRINTS; PR00094; ADENYLTKNASE.
DR   ProDom; PD000657; Adenylate_kin; 1.
DR   TIGRFAMs; TIGR01351; adk; 1.
DR   PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW   Transferase; Kinase; ATP-binding; Complete proteome.
FT   NP_BIND      7     15       ATP (BY SIMILARITY).
FT   CONFLICT    149    149      A -> R (IN REF. 1; CAB11460).
SQ   SEQUENCE   181 AA;  20111 MW;  3023A5EFC359A614 CRC64;

  Query Match             84.4%;  Score 27;  DB 1;  Length 181;
  Best Local Similarity   83.3%;  Pred. No. 50;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||:|||
Db        132 DTDDVV 137


RESULT 11
3HAO_RAT
ID   3HAO_RAT       STANDARD;      PRT;   286 AA.
AC   P46953; P70474; Q64556;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   3-hydroxyanthranilate 3,4-dioxygenase (EC 1.13.11.6) (3-HAO)
DE   (3-hydroxyanthranilic acid dioxygenase) (3-hydroxyanthranilate
DE   oxygenase).
GN   HAAO.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96100955; PubMed=8541664;
RA   Nakagawa Y., Asai H., Mori H., Kitoh J., Nakano K.;
RT   "Increase in the level of mRNA for 3-hydroxyanthranilate 3,4-
RT   dioxygenase in brain of epilepsy-prone El mice.";
RL   Biosci. Biotechnol. Biochem. 59:2191-2192(1995).
RN   [2]
RP   PARTIAL SEQUENCE OF 44-236 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=94245687; PubMed=7514594;
RA   Malherbe P., Kohler C., da Prada M., Lang G., Kiefer V.,
RA   Schwarcz R., Lahm H., Cesura A.M.;
RT   "Molecular cloning and functional expression of human 3-
RT   hydroxyanthranilic-acid dioxygenase.";
RL   J. Biol. Chem. 269:13792-13797(1994).
CC   -!- FUNCTION: CATALYZES THE SYNTHESIS OF THE EXCITOTOXIN QUINOLINIC
CC       ACID (QUIN) FROM 3-HYDROXYANTHRANILIC ACID. THE DIRECT PRODUCT
CC       OF THE REACTION SPONTANEOUSLY REARRANGE TO QUIN.
CC   -!- CATALYTIC ACTIVITY: 3-hydroxyanthranilate + O(2) = 2-amino-3-
CC       carboxymuconate semialdehyde.
CC   -!- COFACTOR: Ferrous ion.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D44494; BAA07937.1; -.
DR   EMBL; D28339; BAA21019.1; -.
DR   InterPro; IPR007113; Cupin_sup.
KW   Oxidoreductase; Dioxygenase; Iron.
FT   CONFLICT     44     44       K -> S (IN REF. 2).
FT   CONFLICT    199    199       S -> F (IN REF. 2).
FT   CONFLICT    204    204       G -> C (IN REF. 2).
FT   CONFLICT    214    214       H -> Y (IN REF. 2).
FT   CONFLICT    229    229       W -> P (IN REF. 2).
SQ   SEQUENCE   286 AA;  32582 MW;  B4F535AD8949DAB7 CRC64;

  Query Match             84.4%;  Score 27;  DB 1;  Length 286;
  Best Local Similarity   83.3%;  Pred. No. 82;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              |||||:
Db        122 DTEDVL 127


RESULT 12
RESA_PLAFP
ID   RESA_PLAFP     STANDARD;      PRT;   304 AA.
AC   Q26005;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ring-infected erythrocyte surface antigen (Fragment).
GN   RESA.
OS   Plasmodium falciparum (isolate Palo Alto / Uganda).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=57270;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94265898; PubMed=8206141;
RA   Kun J., Leet M., Anthony R.L., Kun J.E., Anders R.F.;
RT   "Plasmodium falciparum: a region of polymorphism in the 3' end of the
RT   gene for the ring-infected erythrocyte surface antigen.";
RL   Exp. Parasitol. 78:418-421(1994).
```

```
CC   -!- FUNCTION: RESA MAY DISRUPT THE NORMAL INTERMOLECULAR INTERACTIONS
CC       OF THE CYTOPLASMIC DOMAIN OF BAND 3 AND THEREBY FACILITATE THE
CC       INVAGINATION OF THE RED CELL MEMBRANE WHICH IS NECESSARY FOR THE
CC       FORMATION OF THE PARASITOPHOROUS VACUOLE.
CC   -!- SUBCELLULAR LOCATION: PROBABLY LOCATED ON THE CYTOPLASMIC FACE OF
CC       THE MEMBRANE WHERE IT ASSOCIATES WITH COMPONENTS OF THE MEMBRANE
CC       SKELETON.
CC   -!- SIMILARITY: THE N-TERMINAL SEQUENCE OF BAND 3 SHOW HOMOLOGY WITH
CC       THE REPEAT SEQUENCES OF RESA.
CC   -!- SIMILARITY: Contains 1 J domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55124; CAA38918.1; -.
DR   PIR; S21342; S21342.
DR   InterPro; IPR001623; DnaJ_N.
DR   PROSITE; PS00636; DNAJ_1; PARTIAL.
DR   PROSITE; PS50076; DNAJ_2; PARTIAL.
KW   Malaria; Antigen; Glycoprotein; Repeat.
FT   NON_TER       1      1
FT   CARBOHYD     18     18       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     304    304
SQ   SEQUENCE   304 AA;  35225 MW;  54807B86C85AD75C CRC64;

  Query Match             84.4%;  Score 27;  DB 1;  Length 304;
  Best Local Similarity   83.3%;  Pred. No. 88;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||:|||
Db        103 DTQDVV 108


RESULT 13
KG62_MOUSE
ID   KG62_MOUSE     STANDARD;      PRT;   335 AA.
AC   Q8C0L0;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Thioredoxin-like protein KIAA1162 homolog precursor.
GN   KIAA1162.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- SIMILARITY: Contains 1 thioredoxin domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK030696; BAC27085.1; -.
DR   MGD; MGI:106558; D2Bwg1356e.
DR   InterPro; IPR006662; Thiored.
DR   InterPro; IPR006663; Thioredox_dom2.
DR   Pfam; PF00085; thiored; 1.
DR   PROSITE; PS00194; THIOREDOXIN; 1.
KW   Redox-active center; Electron transport; Transmembrane; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    335       THIOREDOXIN-LIKE PROTEIN KIAA1162
FT                                HOMOLOG.
FT   TRANSMEM    186    206       POTENTIAL.
FT   DOMAIN       26    133       THIOREDOXIN.
FT   DOMAIN      220    288       GLU-RICH.
FT   DISULFID     60     63       REDOX-ACTIVE (POTENTIAL).
SQ   SEQUENCE   335 AA;  37161 MW;  5EE38B2B8C0FBA12 CRC64;

  Query Match             84.4%;  Score 27;  DB 1;  Length 335;
  Best Local Similarity   83.3%;  Pred. No. 98;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||:|||
Db        314 DTQDVV 319


RESULT 14
ZP3_CANFA
ID   ZP3_CANFA      STANDARD;      PRT;   426 AA.
AC   P48831;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Sperm receptor) (Zona pellucida protein C).
GN   ZP3 OR ZPC.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=95143578; PubMed=7841460;
RA   Harris J.D., Hibler D.W., Fontenot G.K., Hsu K.T., Yurewicz E.C.,
RA   Sacco A.G.;
RT   "Cloning and characterization of zona pellucida genes and cDNAs from
RT   a variety of mammalian species: the ZPA, ZPB and ZPC gene families.";
RL   DNA Seq. 4:361-393(1994).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Okazaki Y., Sugimoto M.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination (By similarity).
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides (By
CC       similarity).
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U05780; AAA74387.1; -.
DR   EMBL; D45070; BAA08098.1; -.
DR   PIR; S70396; S70396.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    426       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    385       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    386    406       POTENTIAL.
FT   DOMAIN      407    426       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       43    305       ZP.
FT   CARBOHYD    123    123       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    244    244       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    227    227       L -> P (IN REF. 2).
FT   CONFLICT    307    307       W -> S (IN REF. 2).
FT   CONFLICT    343    343       K -> R (IN REF. 2).
SQ   SEQUENCE   426 AA;  47367 MW;  BE5825A949DCA172 CRC64;

  Query Match             84.4%;  Score 27;  DB 1;  Length 426;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||:|||
Db         83 DTDDVV 88


RESULT 15
YM72_YEAST
ID   YM72_YEAST     STANDARD;      PRT;   590 AA.
AC   Q05021;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 67.6 kDa protein in MRPL44-MTF1 intergenic region.
GN   YMR227C OR YM9959.09C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313268; PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49939; CAA90198.1; -.
DR   PIR; S57594; S57594.
DR   GermOnline; 142902; -.
DR   TRANSFAC; T03088; -.
DR   SGD; S0004840; TAF67.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0000114; P:G1-specific transcription in mitotic cell c. . .; IPI.
DR   InterPro; IPR006751; TAFII55_N.
DR   Pfam; PF04658; TAFII55_N; 1.
KW   Hypothetical protein; Coiled coil.
FT   DOMAIN       50     53       POLY-LYS.
FT   DOMAIN      200    203       POLY-GLU.
FT   DOMAIN      368    373       POLY-GLU.
FT   DOMAIN      413    421       POLY-ASP.
FT   DOMAIN      517    531       POLY-GLU.
FT   DOMAIN      427    549       COILED COIL (POTENTIAL).
SQ   SEQUENCE   590 AA;  67555 MW;  C014E7419B0B1C61 CRC64;

  Query Match             84.4%;  Score 27;  DB 1;  Length 590;
  Best Local Similarity   83.3%;  Pred. No. 1.8e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        211 DTEDLV 216


Search completed: March 31, 2004, 16:46:12
Job time : 8.46667 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:45:43 ; Search time 34.4 Seconds
(without alignments)
60.852 Million cell updates/sec

Title: US-09-909-077-1
Perfect score: 32
Sequence: 1 DTEDVVXX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	93.8	7	9	US-09-777-785-1	Sequence 1, Appli
2	30	93.8	7	9	US-09-777-785-3	Sequence 3, Appli
3	30	93.8	8	10	US-09-909-062-125	Sequence 125, App
4	30	93.8	8	10	US-09-909-062-127	Sequence 127, App
5	30	93.8	8	10	US-09-909-062-128	Sequence 128, App
6	30	93.8	14	9	US-09-747-419-31	Sequence 31, Appl
7	30	93.8	14	14	US-10-259-275-31	Sequence 31, Appl
8	30	93.8	18	14	US-10-300-757-10	Sequence 10, Appl
9	30	93.8	20	10	US-09-775-052-50	Sequence 50, Appl
10	30	93.8	20	12	US-10-232-884-19	Sequence 19, Appl
11	30	93.8	94	14	US-10-300-757-22	Sequence 22, Appl
12	30	93.8	104	9	US-09-864-761-39527	Sequence 39527, A
13	30	93.8	259	12	US-10-424-599-156429	Sequence 156429,
14	30	93.8	505	15	US-10-365-620-62	Sequence 62, Appl
15	30	93.8	677	12	US-10-425-114-55538	Sequence 55538, A
16	30	93.8	729	15	US-10-365-620-64	Sequence 64, Appl
17	30	93.8	764	14	US-10-300-757-20	Sequence 20, Appl
18	30	93.8	1040	9	US-09-929-955-9	Sequence 9, Appli
19	30	93.8	1040	13	US-10-104-966-9	Sequence 9, Appli
20	30	93.8	1070	12	US-10-425-114-57815	Sequence 57815, A
21	30	93.8	3011	9	US-09-742-659-4	Sequence 4, Appli
22	30	93.8	3011	9	US-09-238-076-20	Sequence 20, Appl
23	30	93.8	3011	9	US-09-952-572-9	Sequence 9, Appli
24	30	93.8	3011	9	US-09-929-955-1	Sequence 1, Appli
25	30	93.8	3011	9	US-09-747-419-20	Sequence 20, Appl
26	30	93.8	3011	10	US-09-891-894-3	Sequence 3, Appli
27	30	93.8	3011	10	US-09-995-937-20	Sequence 20, Appl
28	30	93.8	3011	10	US-09-917-563-20	Sequence 20, Appl
29	30	93.8	3011	12	US-10-189-359-14	Sequence 14, Appl
30	30	93.8	3011	13	US-10-104-966-1	Sequence 1, Appli
31	30	93.8	3011	14	US-10-259-275-20	Sequence 20, Appl
32	30	93.8	3011	14	US-10-184-150-3	Sequence 3, Appli
33	30	93.8	3011	15	US-10-328-997-3	Sequence 3, Appli
34	30	93.8	3012	9	US-09-238-076-2	Sequence 2, Appli
35	30	93.8	3012	10	US-09-995-937-2	Sequence 2, Appli
36	30	93.8	3012	10	US-09-917-563-2	Sequence 2, Appli
37	29	90.6	35	9	US-09-864-761-41273	Sequence 41273, A
38	29	90.6	280	14	US-10-156-424A-13	Sequence 13, Appl
39	29	90.6	317	14	US-10-156-761-10897	Sequence 10897, A
40	29	90.6	425	15	US-10-094-749-2593	Sequence 2593, Ap
41	29	90.6	455	12	US-10-282-122A-71740	Sequence 71740, A
42	29	90.6	459	12	US-10-282-122A-70841	Sequence 70841, A
43	29	90.6	560	12	US-10-424-599-216864	Sequence 216864,
44	29	90.6	564	12	US-10-425-114-55010	Sequence 55010, A
45	29	90.6	1077	15	US-10-104-047-2291	Sequence 2291, Ap

ALIGNMENTS

RESULT 1
US-09-777-785-1
; Sequence 1, Application US/09777785
; Patent No. US20020103135A1
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; TITLE OF INVENTION: Azapeptides Useful In The Treatment Of Hepatitis C
; FILE REFERENCE: IN01130K1 US
; CURRENT APPLICATION NUMBER: US/09/777,785
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/181,017
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:azapeptide
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: UNSURE
; LOCATION: (7)
; OTHER INFORMATION: 2[(4-nitrophenoxy)carbonyl]-2-propylhydrazine
US-09-777-785-1

Query Match 93.8%; Score 30; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6
```

RESULT 2

```
US-09-777-785-3
; Sequence 3, Application US/09777785
; Patent No. US20020103135A1
; GENERAL INFORMATION:
;  APPLICANT: Zhang, Rumin
;  TITLE OF INVENTION: Azapeptides Useful In The Treatment Of Hepatitis C
;  FILE REFERENCE: IN01130K1 US
;  CURRENT APPLICATION NUMBER: US/09/777,785
;  CURRENT FILING DATE:  2001-02-06
;  PRIOR APPLICATION NUMBER: 60/181,017
;  PRIOR FILING DATE: 2000-02-08
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:azapeptide
;   NAME/KEY: MOD_RES
;   LOCATION: (1)
;   OTHER INFORMATION: ACETYLATION
;   NAME/KEY: UNSURE
;   LOCATION: (7)
;   OTHER INFORMATION: 2-[(1,2,2,2-tetrachloroethoxy)carbonyl]-2-propylhy
;   OTHER INFORMATION: drazine
US-09-777-785-3

  Query Match             93.8%;  Score 30;  DB 9;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 3
US-09-909-062-125
; Sequence 125, Application US/09909062
; Publication No. US20030036501A1
; GENERAL INFORMATION:
;  APPLICANT: Saksena, Anil K
;  APPLICANT:  Girijavaliabhan, Viyyor M
;  APPLICANT:  Lovey, Raymond G
;  APPLICANT:  Jao, Edwin
;  APPLICANT:  Bennett, Frank
;  APPLICANT:  McCormick, Jinping L
;  APPLICANT:  Pike, Russell E
;  APPLICANT:  Bogen, Stephane L
;  APPLICANT:  Liu, Yi-Tsung
;  APPLICANT:  Arasappan, Ashok
;  APPLICANT:  Pinto, Patrick A
;  APPLICANT:  Njoroge, F George
;  APPLICANT:  Ganguly, Ashit
;  TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C V
;  FILE REFERENCE: IN01157K-US
;  CURRENT APPLICATION NUMBER: US/09/909,062
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: 60/220,109
;  PRIOR FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 149
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: NS5A-NS5B junction sequence
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: ACETYLATION
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: alanine or proline
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (8)..(8)
;   OTHER INFORMATION: norvaline
US-09-909-062-125

  Query Match             93.8%;  Score 30;  DB 10;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 4
US-09-909-062-127
; Sequence 127, Application US/09909062
; Publication No. US20030036501A1
; GENERAL INFORMATION:
;  APPLICANT: Saksena, Anil K
;  APPLICANT:  Girijavaliabhan, Viyyor M
;  APPLICANT:  Lovey, Raymond G
;  APPLICANT:  Jao, Edwin
;  APPLICANT:  Bennett, Frank
;  APPLICANT:  McCormick, Jinping L
;  APPLICANT:  Pike, Russell E
;  APPLICANT:  Bogen, Stephane L
;  APPLICANT:  Liu, Yi-Tsung
;  APPLICANT:  Arasappan, Ashok
;  APPLICANT:  Pinto, Patrick A
;  APPLICANT:  Njoroge, F George
;  APPLICANT:  Ganguly, Ashit
;  TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C V
;  FILE REFERENCE: IN01157K-US
;  CURRENT APPLICATION NUMBER: US/09/909,062
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: 60/220,109
;  PRIOR FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 149
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 127
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: ACETYLATION
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (8)..(8)
;   OTHER INFORMATION: norvaline
US-09-909-062-127

  Query Match             93.8%;  Score 30;  DB 10;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 5
US-09-909-062-128
; Sequence 128, Application US/09909062
```

```
; Publication No. US20030036501A1
; GENERAL INFORMATION:
;  APPLICANT: Saksena, Anil K
;  APPLICANT:  Girijavaliabhan, Viyyor M
;  APPLICANT:  Lovey, Raymond G
;  APPLICANT:  Jao, Edwin
;  APPLICANT:  Bennett, Frank
;  APPLICANT:  McCormick, Jinping L
;  APPLICANT:  Pike, Russell E
;  APPLICANT:  Bogen, Stephane L
;  APPLICANT:  Liu, Yi-Tsung
;  APPLICANT:  Arasappan, Ashok
;  APPLICANT:  Pinto, Patrick A
;  APPLICANT:  Njoroge, F George
;  APPLICANT:  Ganguly, Ashit
;  TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C V
;  FILE REFERENCE: IN01157K-US
;  CURRENT APPLICATION NUMBER: US/09/909,062
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: 60/220,109
;  PRIOR FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 149
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 128
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: ACETYLATION
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (8)..(8)
;   OTHER INFORMATION: norvaline
US-09-909-062-128

  Query Match             93.8%;  Score 30;  DB 10;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 6
US-09-747-419-31
; Sequence 31, Application US/09747419
; Patent No. US20020155582A1
; GENERAL INFORMATION:
;  APPLICANT: Lemon, Stanley
;  APPLICANT:  Yi, MinKyung
;  TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
;  FILE REFERENCE: 265.0007 0101
;  CURRENT APPLICATION NUMBER: US/09/747,419
;  CURRENT FILING DATE:  2000-12-23
;  PRIOR APPLICATION NUMBER: US 60/171,909
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: NS3 recognition site
US-09-747-419-31

  Query Match             93.8%;  Score 30;  DB 9;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 7
US-10-259-275-31
; Sequence 31, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
;  APPLICANT: Lemon, Stanley M.
;  APPLICANT:  Yi, MinKyung
;  TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
;  FILE REFERENCE: 265.0007 0120
;  CURRENT APPLICATION NUMBER: US/10/259,275
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: US 60/171,909
;  PRIOR FILING DATE: 1999-12-23
;  PRIOR APPLICATION NUMBER: US 09/747,419
;  PRIOR FILING DATE: 2000-12-23
;  PRIOR APPLICATION NUMBER: US 60/325,236
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/338,123
;  PRIOR FILING DATE: 2001-11-13
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: NS3 recognition site
US-10-259-275-31

  Query Match             93.8%;  Score 30;  DB 14;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 8
US-10-300-757-10
; Sequence 10, Application US/10300757
; Publication No. US20030083467A1
; GENERAL INFORMATION:
;  APPLICANT: Hoock, Thomas
;  APPLICANT:  Germann, Ursula
;  APPLICANT:  Kwong, Ann
;  TITLE OF INVENTION: FUSION PROTEINS, DNA MOLECULES, VECTORS, AND HOST CELLS
;  TITLE OF INVENTION:  USEFUL FOR MEASURING PROTEASE ACTIVITY
;  FILE REFERENCE: VPI/98-08
;  CURRENT APPLICATION NUMBER: US/10/300,757
;  CURRENT FILING DATE:  2002-11-20
;  PRIOR APPLICATION NUMBER: US/09/570,267
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/144,759
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 18
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-10-300-757-10

  Query Match             93.8%;  Score 30;  DB 14;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 9
US-09-775-052-50
; Sequence 50, Application US/09775052
; Publication No. US20030054000A1
; GENERAL INFORMATION:
;  APPLICANT: Dowdy, Steven F.
;  TITLE OF INVENTION: ANTI-PATHOGEN SYSTEM AND METHODS OF USE THEREOF
;  FILE REFERENCE: 48881/1742
;  CURRENT APPLICATION NUMBER: US/09/775,052
;  CURRENT FILING DATE:  2001-02-01
;  PRIOR APPLICATION NUMBER: 09/208,966
;  PRIOR FILING DATE: 1998-12-10
;  PRIOR APPLICATION NUMBER: 60/082,402
;  PRIOR FILING DATE: 1998-04-20
;  PRIOR APPLICATION NUMBER: 60/069,012
;  PRIOR FILING DATE: 1997-12-10
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: human
US-09-775-052-50

  Query Match             93.8%;  Score 30;  DB 10;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 10
US-10-232-884-19
; Sequence 19, Application US/10232884
; Publication No. US20040043949A1
; GENERAL INFORMATION:
;  APPLICANT: Richardson, Chris
;  APPLICANT:  Kneteman, No. US20040043949A1man
;  APPLICANT:  Hsu, Eric
;  APPLICANT:  Turrell, David Lorne
;  TITLE OF INVENTION: Therapeutic System Targeting Pathogen
;  TITLE OF INVENTION:  Proteases And Uses Thereof
;  FILE REFERENCE: UALB-003
;  CURRENT APPLICATION NUMBER: US/10/232,884
;  CURRENT FILING DATE:  2002-08-30
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
;   FEATURE:
;   OTHER INFORMATION: HCV protease cleavage site
US-10-232-884-19

  Query Match             93.8%;  Score 30;  DB 12;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 11
US-10-300-757-22
; Sequence 22, Application US/10300757
; Publication No. US20030083467A1
; GENERAL INFORMATION:
;  APPLICANT: Hoock, Thomas
;  APPLICANT:  Germann, Ursula
;  APPLICANT:  Kwong, Ann
;  TITLE OF INVENTION: FUSION PROTEINS, DNA MOLECULES, VECTORS, AND HOST CELLS
;  TITLE OF INVENTION:  USEFUL FOR MEASURING PROTEASE ACTIVITY
;  FILE REFERENCE: VPI/98-08
;  CURRENT APPLICATION NUMBER: US/10/300,757
;  CURRENT FILING DATE:  2002-11-20
;  PRIOR APPLICATION NUMBER: US/09/570,267
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/144,759
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 94
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:man-made
;   OTHER INFORMATION: artificial sequence
US-10-300-757-22

  Query Match             93.8%;  Score 30;  DB 14;  Length 94;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         87 DTEDVV 92


RESULT 12
US-09-864-761-39527
; Sequence 39527, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39527
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005522.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.8
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
;   OTHER INFORMATION: SWISSPROT HIT: Q06633, EVALUE 3.00e-57
US-09-864-761-39527

  Query Match             93.8%;  Score 30;  DB 9;  Length 104;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 13
US-10-424-599-156429
; Sequence 156429, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156429
;   LENGTH: 259
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_112277C.1.pep
US-10-424-599-156429

  Query Match             93.8%;  Score 30;  DB 12;  Length 259;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         60 DTEDVV 65


RESULT 14
US-10-365-620-62
; Sequence 62, Application US/10365620
; Publication No. US20040001853A1
; GENERAL INFORMATION:
;  APPLICANT: George, Rajan
;  APPLICANT:  Tyrrell, Lorne
;  APPLICANT:  No. US20040001853A1jaim, Antoine
;  TITLE OF INVENTION: Chimeric Antigens for Eliciting An Immune Response
;  FILE REFERENCE: 656.0016
;  CURRENT APPLICATION NUMBER: US/10/365,620
;  CURRENT FILING DATE:  2003-02-13
;  PRIOR APPLICATION NUMBER: US60/423,578
;  PRIOR FILING DATE: 2003-11-05
;  PRIOR APPLICATION NUMBER: 60/390,564
;  PRIOR FILING DATE: 2002-06-20
;  NUMBER OF SEQ ID NOS: 76
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 62
;   LENGTH: 505
;   TYPE: PRT
;   ORGANISM: ORF of HCV NS5A protein
US-10-365-620-62

  Query Match             93.8%;  Score 30;  DB 15;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 5.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        471 DTEDVV 476


RESULT 15
US-10-425-114-55538
; Sequence 55538, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Asscciated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55538
;   LENGTH: 677
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-GMFLMINSOY106C11_FLI.pep
US-10-425-114-55538

  Query Match             93.8%;  Score 30;  DB 12;  Length 677;
  Best Local Similarity   100.0%;  Pred. No. 7.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         34 DTEDVV 39


Search completed: March 31, 2004, 16:52:55
Job time : 35.4 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:01 ; Search time 50.6667 Seconds
(without alignments)
44.613 Million cell updates/sec

Title: US-09-909-077-1
Perfect score: 32
Sequence: 1 DTEDVVXX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	93.8	7	4	AAG66392	Aag66392 Azapeptid
2	30	93.8	7	5	AAM51805	Aam51805 HCV prote
3	30	93.8	7	5	AAM51807	Aam51807 HCV prote
4	30	93.8	7	5	AAM51806	Aam51806 HCV prote
5	30	93.8	7	6	ABR61795	Abr61795 HCV prote
6	30	93.8	8	4	AAE10046	Aae10046 Hepatitis
7	30	93.8	8	4	AAE10053	Aae10053 Hepatitis
8	30	93.8	8	4	AAE10055	Aae10055 Hepatitis
9	30	93.8	8	4	AAE10048	Aae10048 Hepatitis
10	30	93.8	8	4	AAE10057	Aae10057 Hepatitis
11	30	93.8	8	4	AAE10051	Aae10051 Hepatitis
12	30	93.8	8	4	AAE10054	Aae10054 Hepatitis
13	30	93.8	8	4	AAE10056	Aae10056 Hepatitis
14	30	93.8	8	4	AAE10049	Aae10049 Hepatitis
15	30	93.8	8	5	AAU76968	Aau76968 Hepatitis
16	30	93.8	8	5	ABB07111	Abb07111 Hepatitis
17	30	93.8	8	5	ABB07110	Abb07110 Hepatitis
18	30	93.8	8	5	ABB07108	Abb07108 Hepatitis
19	30	93.8	10	4	AAE10047	Aae10047 Hepatitis
20	30	93.8	10	4	AAE10058	Aae10058 Hepatitis
21	30	93.8	10	4	AAE10052	Aae10052 Hepatitis
22	30	93.8	10	4	AAE10044	Aae10044 Hepatitis
23	30	93.8	12	4	AAE10045	Aae10045 Hepatitis
24	30	93.8	12	4	AAE10050	Aae10050 Hepatitis
25	30	93.8	13	4	AAE10061	Aae10061 Hepatitis
26	30	93.8	13	4	AAE10059	Aae10059 Hepatitis
27	30	93.8	13	4	AAE10062	Aae10062 Hepatitis
28	30	93.8	14	4	AAE10060	Aae10060 Hepatitis
29	30	93.8	14	5	AAM48243	Aam48243 Hepatitis
30	30	93.8	14	6	ABG73198	Abg73198 MKO-Z NS3
31	30	93.8	14	7	ADD67966	Add67966 NS3 recog
32	30	93.8	16	2	AAW09251	Aaw09251 HCV NS3 p
33	30	93.8	16	2	AAW12958	Aaw12958 HCV mutan
34	30	93.8	16	2	AAW01653	Aaw01653 Mutant 5A
35	30	93.8	16	2	AAW47149	Aaw47149 Hepatitis
36	30	93.8	16	2	AAY17895	Aay17895 5A/5B (P8
37	30	93.8	16	3	AAY57208	Aay57208 Native NS
38	30	93.8	17	2	AAW12957	Aaw12957 HCV solub
39	30	93.8	17	2	AAW12959	Aaw12959 HCV mutan
40	30	93.8	17	2	AAW04561	Aaw04561 HCV prote
41	30	93.8	17	2	AAW04574	Aaw04574 HCV NS3 p
42	30	93.8	17	2	AAW01652	Aaw01652 Soluble 5
43	30	93.8	17	2	AAW01654	Aaw01654 Mutant so
44	30	93.8	17	2	AAW47146	Aaw47146 Hepatitis
45	30	93.8	17	2	AAY17896	Aay17896 5A/5B sub

ALIGNMENTS

RESULT 1
AAG66392
ID AAG66392 standard; peptide; 7 AA.
XX
AC AAG66392;
XX
DT 15-OCT-2001 (first entry)
XX
DE Azapeptide #1 useful for treating hepatitis C infection.
XX
KW Virucide; hepatotropic; azapeptide; Hepatitis C viral infection;
KW serine protease inhibitor; antiviral.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 7
FT /note= "Optionally: proline 2-[(4-nitrophenoxy)carbonyl]-
FT 2-propylhydrazine, or:proline 2-[1,2,2,2-
FT tetrachloroethoxy)- carbonyl]-2-propylhydrazine"
XX
PN WO200158929-A1.
XX
PD 16-AUG-2001.
XX
PF 06-FEB-2001; 2001WO-US003768.
XX
PR 08-FEB-2000; 2000US-0181017P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R;
XX
DR WPI; 2001-536524/59.
XX
PT New azapeptide compounds, are inhibitors of hepatitis C virus serine
PT protease, useful in treating hepatitis C infection.
XX
PS Claim 16; Page 35; 44pp; English.
XX
CC The present invention relates to novel azapeptide compounds. The present
CC sequence is one such peptide compound. This peptide compound is a
CC Hepatitis C virus serine protease inhibitor. This peptide compound can be
CC used in the treatment of hepatitis C viral infections, by administrating
CC the present peptide compound, optionally in combination with an antiviral
CC agent and/or an enzyme inhibitor

```
XX
SQ   Sequence 7 AA;

  Query Match            93.8%;  Score 30;  DB 4;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 2
AAM51805
ID   AAM51805 standard; peptide; 7 AA.
XX
AC   AAM51805;
XX
DT   22-JAN-2002  (first entry)
XX
DE   HCV protease inhibition assay substrate peptide #1.
XX
KW   HCV; Hepatitis C virus; virucide; hepatotropic; antiinflammatory;
KW   Hepatitis C; NS3/NS4a serine protease.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /label= OTHER
FT                   /note= "N-terminal acetyl"
FT   Modified-site   7
FT                   /label= OTHER
FT                   /note= "Ala or Pro, modified by Nva"
XX
PN   WO200177113-A2.
XX
PD   18-OCT-2001.
XX
PF   03-APR-2001; 2001WO-US010869.
XX
PR   05-APR-2000; 2000US-0194607P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Chen KX,  Arasappan A,  Venkatraman S,  Parekh TN,  Gu H,  Njoroge FG;
PI   Girijavallabhan VM,  Ganguly A,  Saksena A,  Jao E,  Yao NH,  Prongay AJ;
PI   Madison VS,  Vibulbhan B;
XX
DR   WPI; 2002-017438/02.
XX
PT   New macrocyclic compounds are hepatitis C virus inhibitors (HCV),
PT   especially HCV NS3/NS4a serine protease inhibitors, useful for treating
PT   hepatitis C and related disorders.
XX
PS   Example 111; Page 356; 402pp; English.
XX
CC   The present invention relates to macrocyclic compounds and their
CC   derivatives, which are capable of acting as Hepatitis C virus (HCV)
CC   inhibitors. They are particularly useful for inhibiting HCV NS3/NS4a
CC   serine protease. The compounds can be used to treat disorders associated
CC   with HCV, including hepatitis C. The present sequence is a peptide
CC   substrate used in a HCV protease inhibition assay in the exemplification
CC   of the invention
XX
SQ   Sequence 7 AA;

  Query Match            93.8%;  Score 30;  DB 5;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 3
AAM51807
ID   AAM51807 standard; peptide; 7 AA.
XX
AC   AAM51807;
XX
DT   22-JAN-2002  (first entry)
XX
DE   HCV protease inhibition assay substrate peptide #3.
XX
KW   HCV; Hepatitis C virus; virucide; hepatotropic; antiinflammatory;
KW   Hepatitis C; NS3/NS4a serine protease.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /label= OTHER
FT                   /note= "N-terminal acetyl"
FT   Modified-site   7
FT                   /label= OTHER
FT                   /note= "modified by Nva"
XX
PN   WO200177113-A2.
XX
PD   18-OCT-2001.
XX
PF   03-APR-2001; 2001WO-US010869.
XX
PR   05-APR-2000; 2000US-0194607P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Chen KX,  Arasappan A,  Venkatraman S,  Parekh TN,  Gu H,  Njoroge FG;
PI   Girijavallabhan VM,  Ganguly A,  Saksena A,  Jao E,  Yao NH,  Prongay AJ;
PI   Madison VS,  Vibulbhan B;
XX
DR   WPI; 2002-017438/02.
XX
PT   New macrocyclic compounds are hepatitis C virus inhibitors (HCV),
PT   especially HCV NS3/NS4a serine protease inhibitors, useful for treating
PT   hepatitis C and related disorders.
XX
PS   Example 111; Page 359; 402pp; English.
XX
CC   The present invention relates to macrocyclic compounds and their
CC   derivatives, which are capable of acting as Hepatitis C virus (HCV)
CC   inhibitors. They are particularly useful for inhibiting HCV NS3/NS4a
CC   serine protease. The compounds can be used to treat disorders associated
CC   with HCV, including hepatitis C. The present sequence is a peptide
CC   substrate used in a HCV protease inhibition assay in the exemplification
CC   of the invention
XX
SQ   Sequence 7 AA;

  Query Match            93.8%;  Score 30;  DB 5;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 4
AAM51806
ID   AAM51806 standard; peptide; 7 AA.
XX
AC   AAM51806;
XX
```

DT 22-JAN-2002 (first entry)
XX
DE HCV protease inhibition assay substrate peptide #2.
XX
KW HCV; Hepatitis C virus; virucide; hepatotropic; antiinflammatory;
KW Hepatitis C; NS3/NS4a serine protease.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "N-terminal acetyl"
FT Modified-site 7
FT /label= OTHER
FT /note= "modified by Nva"
XX
PN WO200177113-A2.
XX
PD 18-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US010869.
XX
PR 05-APR-2000; 2000US-0194607P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Chen KX, Arasappan A, Venkatraman S, Parekh TN, Gu H, Njoroge FG;
PI Girijavallabhan VM, Ganguly A, Saksena A, Jao E, Yao NH, Prongay AJ;
PI Madison VS, Vibulbhan B;
XX
DR WPI; 2002-017438/02.
XX
PT New macrocyclic compounds are hepatitis C virus inhibitors (HCV),
PT especially HCV NS3/NS4a serine protease inhibitors, useful for treating
PT hepatitis C and related disorders.
XX
PS Example 111; Page 359; 402pp; English.
XX
CC The present invention relates to macrocyclic compounds and their
CC derivatives, which are capable of acting as Hepatitis C virus (HCV)
CC inhibitors. They are particularly useful for inhibiting HCV NS3/NS4a
CC serine protease. The compounds can be used to treat disorders associated
CC with HCV, including hepatitis C. The present sequence is a peptide
CC substrate used in a HCV protease inhibition assay in the exemplification
CC of the invention
XX
SQ Sequence 7 AA;

Query Match 93.8%; Score 30; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 5
ABR61795
ID ABR61795 standard; peptide; 7 AA.
XX
AC ABR61795;
XX
DT 12-SEP-2003 (first entry)
XX
DE HCV protein derived peptide substrate.
XX
KW HCV; protease; drug monitoring; therapeutic; retroviral;
KW protease inhibitor.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetylation"
FT Modified-site 7
FT /note= "Pro-(Nva)-O-4-phenylazophenyl ester"
XX
PN WO2003040390-A2.
XX
PD 15-MAY-2003.
XX
PF 08-NOV-2002; 2002WO-EP012631.
XX
PR 08-NOV-2001; 2001US-0331117P.
XX
PA (TIBO-) TIBOTEC PHARM LTD.
XX
PI Gulnik S, Yu B, Erickson JW;
XX
DR WPI; 2003-493269/46.
XX
PT Determining the inhibitory potency of an active ingredient in a
PT biological sample, useful for therapeutic drug monitoring comprises
PT relating the signal determined to a reference standard curve prepared
PT with at least one reference.
XX
PS Disclosure; Page 7; 32pp; English.
XX
CC The invention relates to determining the inhibitory potency of ar. active
CC ingredient in a biological sample. The method involves providing a
CC biological sample, a bioactive molecule and a reagent for the bioactive
CC molecule, which are then added to a container, determining the signal and
CC relating the signal to a reference standard curve prepared with at least
CC one reference. The methods are useful for therapeutic drug monitoring by
CC determining the amount or concentration of protease inhibitors, including
CC retroviral protease inhibitors such as HIV inhibitors. The present
CC sequence represents a substrate containing a peptide backbone derived
CC from HCV proteins
XX
SQ Sequence 7 AA;

Query Match 93.8%; Score 30; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 6
AAE10046
ID AAE10046 standard; peptide; 8 AA.
XX
AC AAE10046;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus (HCV) nitroanilide based chromogenic substrate #3.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /note= "Cys modified with 3, 5-dinitroanilide"
XX
PN US6251583-B1.
XX

PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 6; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 7
AAE10053
ID AAE10053 standard; peptide; 8 AA.
XX
AC AAE10053;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #10.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /label= Nva
FT /note= "Nva-O-7-hydroxy-4-methyl-coumarin"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 8; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 8
AAE10055
ID AAE10055 standard; peptide; 8 AA.
XX
AC AAE10055;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #12.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /label= Nva
FT /note= "Nva-O-4-phenylazophenol"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX

PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 17; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 9
AAE10048
ID AAE10048 standard; peptide; 8 AA.
XX
AC AAE10048;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #5.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /label= Abu
FT /note= "Abu-O-4-nitrophenol"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 8; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 10
AAE10057
ID AAE10057 standard; peptide; 8 AA.
XX
AC AAE10057;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus (HCV) fluorogenic substrate #1.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /note= "Cys modified with 7-amido-4-methylcoumarin"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C

PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 19; Col 9; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC fluorogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6
```

RESULT 11
AAE10051
ID AAE10051 standard; peptide; 8 AA.
XX
AC AAE10051;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #8.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	Modified-site	1
FT		/note= "N-acetyl Asp"
FT	Modified-site	8
FT		/label= Nva
FT		/note= "Nva-O-3-nitrophenol"

XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 17; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6
```

RESULT 12
AAE10054
ID AAE10054 standard; peptide; 8 AA.
XX
AC AAE10054;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #11.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	Modified-site	1
FT		/note= "N-acetyl Asp"
FT	Modified-site	8
FT		/label= Nva
FT		/note= "Nva-O-7-hydroxy-4-methyl-coumarin"

XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 17; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a

CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 13
AAE10056
ID AAE10056 standard; peptide; 8 AA.
XX
AC AAE10056;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #13.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /label= Nva
FT /note= "Nva-O-4-phenylazophenol"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 17; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 30; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1 DTEDVV 6

RESULT 14
AAE10049
ID AAE10049 standard; peptide; 8 AA.
XX
AC AAE10049;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #6.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /label= Nva
FT /note= "Nva-O-4-nitrophenol"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 17; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved

```
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             93.8%;  Score 30;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 15
AAU76968
ID   AAU76968 standard; peptide; 8 AA.
XX
AC   AAU76968;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Hepatitis C virus NS3/NS4a serine protease inhibitor #1.
XX
KW   Hepatitis C virus; serine protease; NS3/NS4a; virucide; antiinflammatory;
KW   hepatotropic; protease inhibitor; imidazolidinone compound.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site    1
FT                    /note= "Acetylated"
FT   Misc-difference  7
FT                    /label= Ala, Pro
FT   Modified-site    8
FT                    /label= Nva
FT                    /note= "Nva-OH or Norvaline-OH"
XX
PN   WO200208198-A2.
XX
PD   31-JAN-2002.
XX
PF   19-JUL-2001; 2001WO-US022828.
XX
PR   21-JUL-2000; 2000US-0220110P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Arasappan A,  Parekh T,  Njoroge FG,  Girijavallabhan VM;
PI   Ganguily AK;
XX
DR   WPI; 2002-217041/27.
XX
PT   New imidazolidinone compounds are hepatitis C virus protease inhibitors,
PT   useful for treating disorders associated with hepatitis C virus.
XX
PS   Example 48-57; Page 67; 88pp; English.
XX
CC   The invention describes novel imidazolidinone compounds and their
CC   enantiomers, stereoisomers, rotamers, tautomers, salts and solvates. The
CC   compounds are inhibitors of Hepatitis C virus (HCV) protease, preferably
CC   NS3/NS4a serine protease and can be used in treating disorders associated
CC   with hepatitis C virus (HCV) and HCV protease. This sequence represents a
CC   competitive inhibitor of HCV serine protease described in the method of
CC   the invention
XX
SQ   Sequence 8 AA;

  Query Match             93.8%;  Score 30;  DB 5;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


Search completed: March 31, 2004, 16:45:31
Job time : 51.6667 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:45:43 ; Search time 34.4 Seconds
(without alignments)
60.852 Million cell updates/sec

Title: US-09-909-077-3
Perfect score: 35
Sequence: 1 DTEDVVAX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	97.1	8	10	US-09-909-062-127	Sequence 127, App
2	33	94.3	317	14	US-10-156-761-10897	Sequence 10897, A
3	31	88.6	235	9	US-09-738-626-5245	Sequence 5245, Ap
4	31	88.6	417	15	US-10-369-493-16134	Sequence 16134, A
5	31	88.6	418	15	US-10-369-493-15382	Sequence 15382, A
6	31	88.6	418	15	US-10-369-493-15750	Sequence 15750, A
7	30	85.7	7	9	US-09-777-785-1	Sequence 1, Appli
8	30	85.7	7	9	US-09-777-785-3	Sequence 3, Appli
9	30	85.7	8	10	US-09-909-062-125	Sequence 125, App
10	30	85.7	8	10	US-09-909-062-128	Sequence 128, App
11	30	85.7	14	9	US-09-747-419-31	Sequence 31, Appl
12	30	85.7	14	14	US-10-259-275-31	Sequence 31, Appl
13	30	85.7	18	14	US-10-300-757-10	Sequence 10, Appl
14	30	85.7	20	10	US-09-775-052-50	Sequence 50, Appl
15	30	85.7	20	12	US-10-232-884-19	Sequence 19, Appl
16	30	85.7	94	14	US-10-300-757-22	Sequence 22, Appl
17	30	85.7	104	9	US-09-864-761-39527	Sequence 39527, A
18	30	85.7	259	12	US-10-424-599-156429	Sequence 156429,
19	30	85.7	305	12	US-10-335-977-7299	Sequence 7299, Ap
20	30	85.7	306	12	US-10-335-977-7298	Sequence 7298, Ap
21	30	85.7	382	12	US-10-425-114-50516	Sequence 50516, A
22	30	85.7	425	15	US-10-094-749-2593	Sequence 2593, Ap
23	30	85.7	505	15	US-10-365-620-62	Sequence 62, Appl
24	30	85.7	585	12	US-10-335-977-7300	Sequence 7300, Ap
25	30	85.7	585	12	US-10-335-977-7301	Sequence 7301, Ap
26	30	85.7	677	12	US-10-425-114-55538	Sequence 55538, A
27	30	85.7	729	15	US-10-365-620-64	Sequence 64, Appl
28	30	85.7	764	14	US-10-300-757-20	Sequence 20, Appl
29	30	85.7	813	15	US-10-369-493-9989	Sequence 9989, Ap
30	30	85.7	828	12	US-10-335-977-7302	Sequence 7302, Ap
31	30	85.7	832	15	US-10-369-493-17055	Sequence 17055, A
32	30	85.7	1040	9	US-09-929-955-9	Sequence 9, Appli
33	30	85.7	1040	13	US-10-104-966-9	Sequence 9, Appli
34	30	85.7	1070	12	US-10-425-114-57815	Sequence 57815, A
35	30	85.7	1077	15	US-10-104-047-2291	Sequence 2291, Ap
36	30	85.7	1115	15	US-10-369-493-4175	Sequence 4175, Ap
37	30	85.7	3011	9	US-09-742-659-4	Sequence 4, Appli
38	30	85.7	3011	9	US-09-238-076-20	Sequence 20, Appl
39	30	85.7	3011	9	US-09-952-572-9	Sequence 9, Appli
40	30	85.7	3011	9	US-09-929-955-1	Sequence 1, Appli
41	30	85.7	3011	9	US-09-747-419-20	Sequence 20, Appl
42	30	85.7	3011	10	US-09-891-894-3	Sequence 3, Appli
43	30	85.7	3011	10	US-09-995-937-20	Sequence 20, Appl
44	30	85.7	3011	10	US-09-917-563-20	Sequence 20, Appl
45	30	85.7	3011	12	US-10-189-359-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-909-062-127
; Sequence 127, Application US/09909062
; Publication No. US20030036501A1
; GENERAL INFORMATION:
; APPLICANT: Saksena, Anil K
; APPLICANT: Girijavaliabhan, Viyyor M
; APPLICANT: Lovey, Raymond G
; APPLICANT: Jao, Edwin
; APPLICANT: Bennett, Frank
; APPLICANT: McCormick, Jinping L
; APPLICANT: Pike, Russell E
; APPLICANT: Bogen, Stephane L
; APPLICANT: Liu, Yi-Tsung
; APPLICANT: Arasappan, Ashok
; APPLICANT: Pinto, Patrick A
; APPLICANT: Njoroge, F George
; APPLICANT: Ganguly, Ashit
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C
; FILE REFERENCE: IN01157K-US
; CURRENT APPLICATION NUMBER: US/09/909,062
; CURRENT FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/220,109
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 127
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: ACETYLATION
; FEATURE:
; NAME/KEY: MISC_FEATURE

; LOCATION: (8)..(8)
; OTHER INFORMATION: norvaline
US-09-909-062-127

  Query Match 97.1%; Score 34; DB 10; Length 8;
  Best Local Similarity 100.0%; Pred. No. 9.6e+05;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
      |||||||
Db 1 DTEDVVA 7

RESULT 2
US-10-156-761-10897
; Sequence 10897, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10897
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10897

  Query Match 94.3%; Score 33; DB 14; Length 317;
  Best Local Similarity 85.7%; Pred. No. 94;
  Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
      ||||:||
Db 16 DTEDIVA 22

RESULT 3
US-09-738-626-5245
; Sequence 5245, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5245
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5245

  Query Match 88.6%; Score 31; DB 9; Length 235;
  Best Local Similarity 85.7%; Pred. No. 1.7e+02;
  Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
      ||||:||
Db 9 DTEDLVA 15

RESULT 4
US-10-369-493-16134
; Sequence 16134, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16134
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16134

  Query Match 88.6%; Score 31; DB 15; Length 417;
  Best Local Similarity 85.7%; Pred. No. 3.2e+02;
  Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
      ||:||||
Db 9 DTDDVVA 15

RESULT 5
US-10-369-493-15382
; Sequence 15382, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15382
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Xanthomonas campestris

US-10-369-493-15382

Query Match 88.6%; Score 31; DB 15; Length 418;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVA 7
              ||:||||
Db         11 DTDDVVA 17
```

RESULT 6
US-10-369-493-15750
; Sequence 15750, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15750
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15750

Query Match 88.6%; Score 31; DB 15; Length 418;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVA 7
              ||:||||
Db         11 DTDDVVA 17
```

RESULT 7
US-09-777-785-1
; Sequence 1, Application US/09777785
; Patent No. US20020103135A1
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; TITLE OF INVENTION: Azapeptides Useful In The Treatment Of Hepatitis C
; FILE REFERENCE: IN01130K1 US
; CURRENT APPLICATION NUMBER: US/09/777,785
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/181,017
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:azapeptide
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: UNSURE
; LOCATION: (7)
; OTHER INFORMATION: 2[(4-nitrophenoxy)carbonyl]-2-propylhydrazine
US-09-777-785-1

Query Match 85.7%; Score 30; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6
```

RESULT 8
US-09-777-785-3
; Sequence 3, Application US/09777785
; Patent No. US20020103135A1
; GENERAL INFORMATION:
; APPLICANT: Zhang, Rumin
; TITLE OF INVENTION: Azapeptides Useful In The Treatment Of Hepatitis C
; FILE REFERENCE: IN01130K1 US
; CURRENT APPLICATION NUMBER: US/09/777,785
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/181,017
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:azapeptide
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: ACETYLATION
; NAME/KEY: UNSURE
; LOCATION: (7)
; OTHER INFORMATION: 2-[(1,2,2,2-tetrachloroethoxy)carbonyl]-2-propylhy
; OTHER INFORMATION: drazine
US-09-777-785-3

Query Match 85.7%; Score 30; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6
```

RESULT 9
US-09-909-062-125
; Sequence 125, Application US/09909062
; Publication No. US20030036501A1
; GENERAL INFORMATION:
; APPLICANT: Saksena, Anil K
; APPLICANT: Girijavaliabhan, Viyyor M
; APPLICANT: Lovey, Raymond G
; APPLICANT: Jao, Edwin
; APPLICANT: Bennett, Frank
; APPLICANT: McCormick, Jinping L
; APPLICANT: Pike, Russell E
; APPLICANT: Bogen, Stephane L
; APPLICANT: Liu, Yi-Tsung
; APPLICANT: Arasappan, Ashok
; APPLICANT: Pinto, Patrick A
; APPLICANT: Njoroge, F George
; APPLICANT: Ganguly, Ashit
; TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C V
; FILE REFERENCE: IN01157K-US
; CURRENT APPLICATION NUMBER: US/09/909,062
; CURRENT FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/220,109
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125

```
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: NS5A-NS5B junction sequence
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: ACETYLATION
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (7)..(7)
;   OTHER INFORMATION: alanine or proline
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (8)..(8)
;   OTHER INFORMATION: norvaline
US-09-909-062-125

  Query Match             85.7%;  Score 30;  DB 10;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 10
US-09-909-062-128
; Sequence 128, Application US/09909062
; Publication No. US20030036501A1
; GENERAL INFORMATION:
;  APPLICANT: Saksena, Anil K
;  APPLICANT:  Girijavaliabhan, Viyyor M
;  APPLICANT:  Lovey, Raymond G
;  APPLICANT:  Jao, Edwin
;  APPLICANT:  Bennett, Frank
;  APPLICANT:  McCormick, Jinping L
;  APPLICANT:  Pike, Russell E
;  APPLICANT:  Bogen, Stephane L
;  APPLICANT:  Liu, Yi-Tsung
;  APPLICANT:  Arasappan, Ashok
;  APPLICANT:  Pinto, Patrick A
;  APPLICANT:  Njoroge, F George
;  APPLICANT:  Ganguly, Ashit
;  TITLE OF INVENTION: NOVEL PEPTIDES AS NS3-SERINE PROTEASE INHIBITORS OF HEPATITIS C V
;  FILE REFERENCE: IN01157K-US
;  CURRENT APPLICATION NUMBER: US/09/909,062
;  CURRENT FILING DATE:  2001-07-19
;  PRIOR APPLICATION NUMBER: 60/220,109
;  PRIOR FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 149
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 128
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: ACETYLATION
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (8)..(8)
;   OTHER INFORMATION: norvaline
US-09-909-062-128

  Query Match             85.7%;  Score 30;  DB 10;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.6e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          1 DTEDVV 6


RESULT 11
US-09-747-419-31
; Sequence 31, Application US/09747419
; Patent No. US20020155582A1
; GENERAL INFORMATION:
;  APPLICANT: Lemon, Stanley
;  APPLICANT:  Yi, MinKyung
;  TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
;  FILE REFERENCE: 265.0007 0101
;  CURRENT APPLICATION NUMBER: US/09/747,419
;  CURRENT FILING DATE:  2000-12-23
;  PRIOR APPLICATION NUMBER: US 60/171,909
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: NS3 recognition site
US-09-747-419-31

  Query Match             85.7%;  Score 30;  DB 9;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 12
US-10-259-275-31
; Sequence 31, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
;  APPLICANT: Lemon, Stanley M.
;  APPLICANT:  Yi, MinKyung
;  TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
;  FILE REFERENCE: 265.0007 0120
;  CURRENT APPLICATION NUMBER: US/10/259,275
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: US 60/171,909
;  PRIOR FILING DATE: 1999-12-23
;  PRIOR APPLICATION NUMBER: US 09/747,419
;  PRIOR FILING DATE: 2000-12-23
;  PRIOR APPLICATION NUMBER: US 60/325,236
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/338,123
;  PRIOR FILING DATE: 2001-11-13
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: NS3 recognition site
US-10-259-275-31

  Query Match             85.7%;  Score 30;  DB 14;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
```

```
Db          3 DTEDVV 8


RESULT 13
US-10-300-757-10
; Sequence 10, Application US/10300757
; Publication No. US20030083467A1
; GENERAL INFORMATION:
;  APPLICANT: Hoock, Thomas
;  APPLICANT:  Germann, Ursula
;  APPLICANT:  Kwong, Ann
;  TITLE OF INVENTION: FUSION PROTEINS, DNA MOLECULES, VECTORS, AND HOST CELLS
;  TITLE OF INVENTION:  USEFUL FOR MEASURING PROTEASE ACTIVITY
;  FILE REFERENCE: VPI/98-08
;  CURRENT APPLICATION NUMBER: US/10/300,757
;  CURRENT FILING DATE:  2002-11-20
;  PRIOR APPLICATION NUMBER: US/09/570,267
;  PRIOR FILING DATE: 2000-05-12
;  PRIOR APPLICATION NUMBER: 09/144,759
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 18
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-10-300-757-10

  Query Match             85.7%;  Score 30;  DB 14;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 14
US-09-775-052-50
; Sequence 50, Application US/09775052
; Publication No. US20030054000A1
; GENERAL INFORMATION:
;  APPLICANT: Dowdy, Steven F.
;  TITLE OF INVENTION: ANTI-PATHOGEN SYSTEM AND METHODS OF USE THEREOF
;  FILE REFERENCE: 48881/1742
;  CURRENT APPLICATION NUMBER: US/09/775,052
;  CURRENT FILING DATE:  2001-02-01
;  PRIOR APPLICATION NUMBER: 09/208,966
;  PRIOR FILING DATE: 1998-12-10
;  PRIOR APPLICATION NUMBER: 60/082,402
;  PRIOR FILING DATE: 1998-04-20
;  PRIOR APPLICATION NUMBER: 60/069,012
;  PRIOR FILING DATE: 1997-12-10
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: human
US-09-775-052-50

  Query Match             85.7%;  Score 30;  DB 10;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


RESULT 15
US-10-232-884-19
; Sequence 19, Application US/10232884
; Publication No. US20040043949A1
; GENERAL INFORMATION:
;  APPLICANT: Richardson, Chris
;  APPLICANT:  Kneteman, No. US20040043949A1man
;  APPLICANT:  Hsu, Eric
;  APPLICANT:  Turrell, David Lorne
;  TITLE OF INVENTION: Therapeutic System Targeting Pathogen
;  TITLE OF INVENTION:  Proteases And Uses Thereof
;  FILE REFERENCE: UALB-003
;  CURRENT APPLICATION NUMBER: US/10/232,884
;  CURRENT FILING DATE:  2002-08-30
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
;   FEATURE:
;   OTHER INFORMATION: HCV protease cleavage site
US-10-232-884-19

  Query Match             85.7%;  Score 30;  DB 12;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db          3 DTEDVV 8


Search completed: March 31, 2004, 16:52:57
Job time : 35.4 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:40:57 ; Search time 25.6 Seconds
(without alignments)
73.950 Million cell updates/sec

Title: US-09-909-077-2
Perfect score: 25
Sequence: 1 DXLIXC 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	92.0	100	17	Q9YAP3	Q9yap3 aeropyrum p
2	23	92.0	154	10	O46320	O46320 gracilaria
3	23	92.0	176	16	Q7UHM2	Q7uhm2 rhodopirell
4	23	92.0	217	16	Q8DT33	Q8dt33 streptococc
5	23	92.0	275	10	Q9CAY0	Q9cay0 arabidopsis
6	23	92.0	300	10	Q8VYK1	Q8vyk1 arabidopsis
7	23	92.0	310	5	O17096	O17096 caenorhabdi
8	23	92.0	319	16	Q9L247	Q91247 streptomyce
9	23	92.0	351	12	Q9WLC8	Q9wlc8 ateline her
10	23	92.0	365	3	Q9P7S4	Q9p7s4 schizosacch
11	23	92.0	367	16	Q89M62	Q89m62 bradyrhizob
12	23	92.0	385	16	Q7UQF0	Q7uqf0 rhodopirell
13	23	92.0	387	16	Q92T60	Q92t60 rhizobium m
14	23	92.0	387	16	Q8UB99	Q8ub99 agrobacteri
15	23	92.0	397	16	Q82WI2	Q82wi2 nitrosomona
16	23	92.0	403	16	Q82XE1	Q82xe1 nitrosomona
17	23	92.0	409	2	Q849P2	Q849p2 pseudomonas
18	23	92.0	409	16	Q88B61	Q88b61 pseudomonas
19	23	92.0	412	5	Q86KN3	Q86kn3 dictyosteli
20	23	92.0	414	16	O53279	O53279 mycobacteri
21	23	92.0	414	16	Q7TXE8	Q7txe8 mycobacteri
22	23	92.0	438	16	Q9CBR3	Q9cbr3 mycobacteri
23	23	92.0	444	2	Q9APH9	Q9aph9 legionella
24	23	92.0	468	16	Q8EST9	Q8est9 oceanobacil
25	23	92.0	476	16	Q8R9M4	Q8r9m4 thermoanaer
26	23	92.0	502	11	Q8BVF9	Q8bvf9 mus musculu
27	23	92.0	530	16	Q890B7	Q890b7 lactobacill
28	23	92.0	540	5	Q7YZF2	Q7yzf2 tetrahymena
29	23	92.0	554	5	Q20370	Q20370 caenorhabdi
30	23	92.0	564	16	Q7U484	Q7u484 synechococc
31	23	92.0	569	10	O64506	O64506 arabidopsis
32	23	92.0	582	3	O60174	O60174 schizosacch
33	23	92.0	603	5	O76354	O76354 caenorhabdi
34	23	92.0	656	16	Q7TZ26	Q7tz26 mycobacteri
35	23	92.0	665	4	Q8IVP2	Q8ivp2 homc sapien
36	23	92.0	678	4	Q86WT1	Q86wt1 homc sapien
37	23	92.0	683	12	Q9YTN7	Q9ytn7 ateline her
38	23	92.0	683	12	Q80BP8	Q80bp8 saimiriine
39	23	92.0	733	11	Q8BLJ4	Q8blj4 mus musculu
40	23	92.0	769	5	Q9W1K9	Q9w1k9 drosophila
41	23	92.0	1118	16	P73197	P73197 synechocyst
42	23	92.0	1207	5	Q9V3I5	Q9v3i5 drosophila
43	23	92.0	1270	13	O12982	O12982 xenopus lae
44	23	92.0	1354	5	Q9VKK1	Q9vkk1 drosophila
45	23	92.0	1357	5	Q8IJR0	Q8ijr0 plasmodium

ALIGNMENTS

RESULT 1
Q9YAP3
ID Q9YAP3 PRELIMINARY; PRT; 100 AA.
AC Q9YAP3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE1900.
GN APE1900.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80905.1; -.
DR PIR; D72577; D72577.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 100 AA; 11001 MW; 55043D1AD0A1D431 CRC64;

Query Match 92.0%; Score 23; DB 17; Length 100;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DXLIXC 6
              | || |
Db         41 DTLITC 46
```

```
RESULT 2
O46320
ID   O46320      PRELIMINARY;      PRT;   154 AA.
AC   O46320;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   ORF1478.
OS   Gracilaria robusta.
OG   Plasmid Gro4059.
OC   Eukaryota; Rhodophyta; Florideophyceae; Gracilariales; Gracilariaceae;
OC   Gracilaria.
OX   NCBI_TaxID=38400;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goff L.J., Moon D.A.;
RT   "Plasmids of the red algae Gracilaria and Gracilariopsis
RT   (Gracilariales): Molecular characterization and cellular
RT   localization.";
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF034717; AAC04723.1; -.
DR   GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW   Plasmid.
SQ   SEQUENCE   154 AA;  17468 MW;  FC49905FCBFAD1FC CRC64;

  Query Match            92.0%;  Score 23;  DB 10;  Length 154;
  Best Local Similarity  66.7%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db         81 DTLISC 86


RESULT 3
Q7UHM2
ID   Q7UHM2      PRELIMINARY;      PRT;   176 AA.
AC   Q7UHM2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to (Di)nucleoside polyphosphate hydrolase (EC 3.6.1.-).
GN   NUDH OR RB13111.
OS   Rhodopirellula baltica.
OC   Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC   Planctomycetaceae; Pirellula.
OX   NCBI_TaxID=117;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1;
RX   MEDLINE=22735913; PubMed=12835416;
RA   Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA   Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA   Schlesner H., Amann R., Reinhardt R.;
RT   "Complete genome sequence of the marine planctomycete Pirellula sp.
RT   strain 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR   EMBL; BX294156; CAD77948.1; -.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   176 AA;  20021 MW;  96F99E3D30672AC5 CRC64;

  Query Match            92.0%;  Score 23;  DB 16;  Length 176;
  Best Local Similarity  66.7%;  Pred. No. 2.6e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        128 DTLITC 133


RESULT 4
Q8DT33
ID   Q8DT33      PRELIMINARY;      PRT;   217 AA.
AC   Q8DT33;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative potassium uptake protein TrkA.
GN   TRK OR SMU.1562.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UA159 / ATCC 700610 / Serotype C;
RX   MEDLINE=22295063; PubMed=12397186;
RA   Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA   Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA   Li S., Zhu H., Najar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT   "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT   pathogen.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR   EMBL; AE014987; AAN59208.1; -.
DR   GO; GO:0008324; F:cation transporter activity; IEA.
DR   GO; GO:0006813; P:potassium ion transport; IEA.
DR   InterPro; IPR006036; TrkA_Kuptake.
DR   InterPro; IPR003148; TrkA_N.
DR   Pfam; PF02254; TrkA-N; 1.
DR   PRINTS; PR00335; KUPTAKETRKA.
KW   Complete proteome.
SQ   SEQUENCE   217 AA;  24206 MW;  495A09836E094F1F CRC64;

  Query Match            92.0%;  Score 23;  DB 16;  Length 217;
  Best Local Similarity  66.7%;  Pred. No. 3.1e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db         67 DSLISC 72


RESULT 5
Q9CAY0
ID   Q9CAY0      PRELIMINARY;      PRT;   275 AA.
AC   Q9CAY0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F24K9.14.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
```

```
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
DR   EMBL; AC008153; AAG51443.1; -.
DR   InterPro; IPR008278; 4-PPT_transf.
DR   Pfam; PF01648; ACPS; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   275 AA;  31284 MW;  1A1055029E1CC638 CRC64;

  Query Match             92.0%;  Score 23;  DB 10;  Length 275;
  Best Local Similarity   66.7%;  Pred. No. 3.9e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        127 DSLIAC 132


RESULT 6
Q8VYK1
ID   Q8VYK1      PRELIMINARY;      PRT;   300 AA.
AC   Q8VYK1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   AT3g11470/F24K9_14.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA   Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RT   "Arabidopsis ORF clones.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY070484; AAL49949.1; -.
DR   EMBL; AY091696; AAM10295.1; -.
DR   InterPro; IPR008278; 4-PPT_transf.
DR   Pfam; PF01648; ACPS; 1.
SQ   SEQUENCE   300 AA;  34288 MW;  8D2280B9296347AF CRC64;

  Query Match             92.0%;  Score 23;  DB 10;  Length 300;
  Best Local Similarity   66.7%;  Pred. No. 4.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        132 DSLIAC 137


RESULT 7
O17096
ID   O17096      PRELIMINARY;      PRT;   310 AA.
AC   O17096;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F40H7.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Pauley A., Le T.T.;
RT   "The sequence of C. elegans cosmid F40H7.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF024498; AAF39802.1; -.
DR   PIR; T32283; T32283.
DR   WormPep; F40H7.2; CE10200.
KW   Hypothetical protein.
SQ   SEQUENCE   310 AA;  34838 MW;  BC36CF5593052CB4 CRC64;

  Query Match             92.0%;  Score 23;  DB 5;  Length 310;
  Best Local Similarity   66.7%;  Pred. No. 4.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        171 DTLITC 176


RESULT 8
Q9L247
ID   Q9L247      PRELIMINARY;      PRT;   319 AA.
AC   Q9L247;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative ABC transporter ATP-binding protein.
GN   SCO2674 OR SC6D10.17.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
```

```
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AL939113; CAB71212.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW   ATP-binding; Transport; Complete proteome.
SQ   SEQUENCE   319 AA;  34351 MW;  F9FB9564C1AEE3B6 CRC64;

  Query Match             92.0%;  Score 23;  DB 16;  Length 319;
  Best Local Similarity   66.7%;  Pred. No. 4.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         26 DALITC 31


RESULT 9
Q9WLC8
ID   Q9WLC8      PRELIMINARY;      PRT;   351 AA.
AC   Q9WLC8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Orf 29B.
OS   Ateline herpesvirus 3.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=85618;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=73;
RX   MEDLINE=20091363; PubMed=10623770;
RA   Albrecht J.C.;
RT   "Primary structure of the Herpesvirus Ateles genome.";
RL   J. Virol. 74:1033-1037(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=73;
RA   Albrecht J.-C., Fleckenstein B.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083424; AAC95555.1; -.
DR   GO; GO:0006323; P:DNA packaging; IEA.
DR   InterPro; IPR003498; DNA_pack_C.
DR   Pfam; PF02499; DNA_pack_C; 1.
SQ   SEQUENCE   351 AA;  38996 MW;  FD3AC95191E8F914 CRC64;

  Query Match             92.0%;  Score 23;  DB 12;  Length 351;
  Best Local Similarity   66.7%;  Pred. No. 4.8e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         53 DSLISC 58


RESULT 10
Q9P7S4
ID   Q9P7S4      PRELIMINARY;      PRT;   365 AA.
AC   Q9P7S4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to yeast mpt1 like-protein, possible involvement in
DE   transcription.
GN   SPAC23G3.09.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RA   Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL138854; CAB72234.1; -.
DR   PIR; T50183; T50183.
DR   GeneDB_SPombe; SPAC23G3.09; -.
DR   InterPro; IPR007900; TAF4.
DR   Pfam; PF05236; TAF4; 1.
SQ   SEQUENCE   365 AA;  40546 MW;  D024E4707701C731 CRC64;

  Query Match             92.0%;  Score 23;  DB 3;  Length 365;
  Best Local Similarity   66.7%;  Pred. No. 5e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db         83 DALISC 88


RESULT 11
Q89M62
ID   Q89M62      PRELIMINARY;      PRT;   367 AA.
AC   Q89M62;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Blr4331 protein.
GN   BLR4331.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005950; BAC49596.1; -.
DR   InterPro; IPR005338; UPF0075.
DR   Pfam; PF03702; UPF0075; 1.
KW   Complete proteome.
SQ   SEQUENCE   367 AA;  38986 MW;  EF4BEC8B1E1ACD2E CRC64;

  Query Match             92.0%;  Score 23;  DB 16;  Length 367;
  Best Local Similarity   66.7%;  Pred. No. 5e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        187 DTLIAC 192


RESULT 12
Q7UQF0
ID   Q7UQF0      PRELIMINARY;      PRT;   385 AA.
AC   Q7UQF0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   RB6365.
OS   Rhodopirellula baltica.
OC   Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC   Planctomycetaceae; Pirellula.
OX   NCBI_TaxID=117;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1;
RX   MEDLINE=22735913; PubMed=12835416;
RA   Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA   Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA   Schlesner H., Amann R., Reinhardt R.;
RT   "Complete genome sequence of the marine planctomycete Pirellula sp.
RT   strain 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR   EMBL; BX294144; CAD74753.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   385 AA;  42939 MW;  AD8683401E089398 CRC64;

  Query Match             92.0%;  Score 23;  DB 16;  Length 385;
  Best Local Similarity   66.7%;  Pred. No. 5.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        231 DALIAC 236


RESULT 13
Q92T60
ID   Q92T60      PRELIMINARY;      PRT;   387 AA.
AC   Q92T60;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein R00120.
GN   R00120 OR SMC04132.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR   EMBL; AL591782; CAC41507.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   387 AA;  42231 MW;  ABDB6BE71F07EC6E CRC64;

  Query Match             92.0%;  Score 23;  DB 16;  Length 387;
  Best Local Similarity   66.7%;  Pred. No. 5.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | ||| |
Db        111 DALIAC 116


RESULT 14
Q8UB99
ID   Q8UB99      PRELIMINARY;      PRT;   387 AA.
AC   Q8UB99;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein Atu3117.
GN   ATU3117 OR AGR_L_3385.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009242; AAL43933.1; -.
DR   EMBL; AE008371; AAK90266.1; -.
DR   PIR; AG2939; AG2939.
DR   PIR; H98342; H98342.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   387 AA;  42620 MW;  946585B673857FB5 CRC64;
```

```
  Query Match             92.0%;  Score 23;  DB 16;  Length 387;
  Best Local Similarity   66.7%;  Pred. No. 5.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        112 DALIAC 117


RESULT 15
Q82WI2
ID   Q82WI2      PRELIMINARY;      PRT;   397 AA.
AC   Q82WI2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Tryptophan synthase, beta chain (EC 4.2.1.20).
GN   TRPB OR NE0693.
OS   Nitrosomonas europaea.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC   Nitrosomonadaceae; Nitrosomonas.
OX   NCBI_TaxID=915;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19718 / IFO 14298;
RX   MEDLINE=22586410; PubMed=12700255;
RA   Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA   Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA   Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT   "Complete genome sequence of the ammonia-oxidizing bacterium and
RT   obligate chemolithoautotroph Nitrosomonas europaea.";
RL   J. Bacteriol. 185:2759-2773(2003).
DR   EMBL; BX321858; CAD84604.1; -.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0004834; F:tryptophan synthase activity; IEA.
DR   GO; GO:0006520; P:amino acid metabolism; IEA.
DR   GO; GO:0006568; P:tryptophan metabolism; IEA.
DR   InterPro; IPR001926; B6_enzyme_beta.
DR   InterPro; IPR006654; Trp_synth_beta.
DR   InterPro; IPR006653; Trp_synth_b_rel.
DR   Pfam; PF00291; PALP; 1.
DR   TIGRFAMs; TIGR00263; trpB; 1.
DR   PROSITE; PS00168; TRP_SYNTHASE_BETA; 1.
KW   Lyase; Complete proteome.
SQ   SEQUENCE   397 AA;  43399 MW;  8BD08F3CA35827B4 CRC64;

  Query Match             92.0%;  Score 23;  DB 16;  Length 397;
  Best Local Similarity   66.7%;  Pred. No. 5.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DXLIXC 6
              | || |
Db        228 DALIAC 233


Search completed: March 31, 2004, 16:48:36
Job time : 28.6 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:01 ; Search time 50.6667 Seconds
(without alignments)
44.613 Million cell updates/sec

Title: US-09-909-077-3
Perfect score: 35
Sequence: 1 DTEDVVAX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	97.1	7	5	AAM51806	Aam51806 HCV prote
2	34	97.1	8	4	AAE10046	Aae10046 Hepatitis
3	34	97.1	8	4	AAE10053	Aae10053 Hepatitis
4	34	97.1	8	4	AAE10055	Aae10055 Hepatitis
5	34	97.1	8	4	AAE10048	Aae10048 Hepatitis
6	34	97.1	8	4	AAE10051	Aae10051 Hepatitis
7	34	97.1	8	4	AAE10049	Aae10049 Hepatitis
8	34	97.1	8	5	ABB07110	Abb07110 Hepatitis
9	34	97.1	10	4	AAE10047	Aae10047 Hepatitis
10	34	97.1	10	4	AAE10058	Aae10058 Hepatitis
11	34	97.1	10	4	AAE10052	Aae10052 Hepatitis
12	34	97.1	12	4	AAE10050	Aae10050 Hepatitis
13	34	97.1	16	2	AAW12958	Aaw12958 HCV mutan
14	34	97.1	16	2	AAW01653	Aaw01653 Mutant 5A
15	34	97.1	17	2	AAW12959	Aaw12959 HCV mutan
16	34	97.1	17	2	AAW04561	Aaw04561 HCV prote
17	34	97.1	17	2	AAW01654	Aaw01654 Mutant so
18	34	97.1	17	2	AAY17896	Aay17896 5A/5B sub
19	34	97.1	17	4	AAE10067	Aae10067 Hepatitis
20	34	97.1	18	2	AAW12961	Aaw12961 HCV mutan
21	34	97.1	18	2	AAW04575	Aaw04575 HCV NS3 p
22	34	97.1	18	2	AAW04562	Aaw04562 HCV prote
23	34	97.1	18	2	AAW01656	Aaw01656 Soluble 5
24	31	88.6	235	4	AAB79243	Aab79243 Corynebac
25	31	88.6	235	4	AAG91491	Aag91491 C glutami
26	31	88.6	275	7	ADD13413	Add13413 C. glutam
27	30	85.7	7	4	AAG66392	Aag66392 Azapeptid
28	30	85.7	7	5	AAM51805	Aam51805 HCV prote
29	30	85.7	7	5	AAM51807	Aam51807 HCV prote
30	30	85.7	7	6	ABR61795	Abr61795 HCV prote
31	30	85.7	8	4	AAE10057	Aae10057 Hepatitis
32	30	85.7	8	4	AAE10054	Aae10054 Hepatitis
33	30	85.7	8	4	AAE10056	Aae10056 Hepatitis
34	30	85.7	8	5	AAU76968	Aau76968 Hepatitis
35	30	85.7	8	5	ABB07111	Abb07111 Hepatitis
36	30	85.7	8	5	ABB07108	Abb07108 Hepatitis
37	30	85.7	10	4	AAE10044	Aae10044 Hepatitis
38	30	85.7	12	4	AAE10045	Aae10045 Hepatitis
39	30	85.7	13	4	AAE10061	Aae10061 Hepatitis
40	30	85.7	13	4	AAE10059	Aae10059 Hepatitis
41	30	85.7	13	4	AAE10062	Aae10062 Hepatitis
42	30	85.7	14	4	AAE10060	Aae10060 Hepatitis
43	30	85.7	14	5	AAM48243	Aam48243 Hepatitis
44	30	85.7	14	6	ABG73198	Abg73198 MKO-Z NS3
45	30	85.7	14	7	ADD67966	Add67966 NS3 recog

ALIGNMENTS

RESULT 1
AAM51806
ID AAM51806 standard; peptide; 7 AA.
XX
AC AAM51806;
XX
DT 22-JAN-2002 (first entry)
XX
DE HCV protease inhibition assay substrate peptide #2.
XX
KW HCV; Hepatitis C virus; virucide; hepatotropic; antiinflammatory;
KW Hepatitis C; NS3/NS4a serine protease.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "N-terminal acetyl"
FT Modified-site 7
FT /label= OTHER
FT /note= "modified by Nva"
XX
PN WO200177113-A2.
XX
PD 18-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US010869.
XX
PR 05-APR-2000; 2000US-0194607P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Chen KX, Arasappan A, Venkatraman S, Parekh TN, Gu H, Njoroge FG;
PI Girijavallabhan VM, Ganguly A, Saksena A, Jao E, Yao NH, Prongay AJ;
PI Madison VS, Vibulbhan B;
XX
DR WPI; 2002-017438/02.
XX
PT New macrocyclic compounds are hepatitis C virus inhibitors (HCV),
PT especially HCV NS3/NS4a serine protease inhibitors, useful for treating
PT hepatitis C and related disorders.
XX
PS Example 111; Page 359; 402pp; English.
XX
CC The present invention relates to macrocyclic compounds and their
CC derivatives, which are capable of acting as Hepatitis C virus (HCV)
CC inhibitors. They are particularly useful for inhibiting HCV NS3/NS4a

CC serine protease. The compounds can be used to treat disorders associated
CC with HCV, including hepatitis C. The present sequence is a peptide
CC substrate used in a HCV protease inhibition assay in the exemplification
CC of the invention
XX
SQ Sequence 7 AA;

Query Match 97.1%; Score 34; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
|||||||
Db 1 DTEDVVA 7

RESULT 2
AAE10046
ID AAE10046 standard; peptide; 8 AA.
XX
AC AAE10046;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus (HCV) nitroanilide based chromogenic substrate #3.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /note= "Cys modified with 3, 5-dinitroanilide"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 6; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

Query Match 97.1%; Score 34; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
|||||||
Db 1 DTEDVVA 7

RESULT 3
AAE10053
ID AAE10053 standard; peptide; 8 AA.
XX
AC AAE10053;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus nitrophenol and ester based chromogenic substrate #10.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Asp"
FT Modified-site 8
FT /label= Nva
FT /note= "Nva-O-7-hydroxy-4-methyl-coumarin"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 8; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 8 AA;

```
  Query Match             97.1%;  Score 34;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 4
AAE10055
ID   AAE10055 standard; peptide; 8 AA.
XX
AC   AAE10055;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #12.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   8
FT                   /label= Nva
FT                   /note= "Nva-O-4-phenylazophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 5
AAE10048
ID   AAE10048 standard; peptide; 8 AA.
XX
AC   AAE10048;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #5.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   8
FT                   /label= Abu
FT                   /note= "Abu-O-4-nitrophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 8; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
```

```
Db          1 DTEDVVA 7


RESULT 6
AAE10051
ID   AAE10051 standard; peptide; 8 AA.
XX
AC   AAE10051;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #8.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   8
FT                   /label= Nva
FT                   /note= "Nva-O-3-nitrophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7
```

```
RESULT 7
AAE10049
ID   AAE10049 standard; peptide; 8 AA.
XX
AC   AAE10049;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #6.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   8
FT                   /label= Nva
FT                   /note= "Nva-O-4-nitrophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 8 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 8
ABB07110
ID   ABB07110 standard; peptide; 8 AA.
```

XX
AC ABB07110;
XX
DT 25-JUN-2002 (first entry)
XX
DE Hepatitis C virus NS3-serine protease inhibitor related peptide #34.
XX
KW Hepatitis C virus; HCV; NS3-serine protease inhibitor; hepatitis;
KW HCV protease inhibitor; infection; virucide; hepatotrophic.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "acetylated"
FT Modified-site 8
FT /label= Nva
FT /note= "norvaline"
XX
PN WO200208256-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US022826.
XX
PR 21-JUL-2000; 2000US-0220109P.
XX
PA (SCHE) SCHERING CORP.
PA (CORV-) CORVAS INT INC.
XX
PI Saksena AK, Girijavallabhan VM, Lovey RG, Jao EE, Bennett F;
PI Mccormick J, Wang H, Pike RE, Bogen SL, Liu Y, Arasappan A;
PI Parekh T, Pinto PA, Njoroge FG, Ganguly AK, Brunck TK, Kemp SJ;
PI Levy OE, Lim-Wilby M;
XX
DR WPI; 2002-361644/39.
XX
PT Novel peptide inhibitor compounds of hepatitis virus NS3/NS4a serine
PT protease, useful for treating hepatitis C virus disorders.
XX
PS Example 9; Page 135; 196pp; English.
XX
CC The present invention describes a peptide compound (I) exhibiting
CC hepatitis C virus (HCV) protease inhibitory activity, including
CC enantiomers, stereoisomers, rotomers and tautomers, pharmaceutically
CC acceptable salts, solvates or derivatives. Also described are: (1) a
CC pharmaceutical composition (II) comprising (I); and (2) preparing (II)
CC for treating disorders associated with HCV protease involving bringing
CC into intimate contact (I) and a carrier. (I) has virucide and
CC hepatotrophic activities and can be used as HCV NS3/NS4a serine protease
CC inhibitors. (I) is useful for manufacturing a medicament to treat
CC disorders associated with HCV protease. (I) can be used for modulating
CC activity of HCV protease preferably, HCV NS3/NS4a protease and for
CC modulating the processing of HCV polypeptide. (II) is useful for treating
CC disorders associated with HCV and for treating disorders associated with
CC HCV protease. (I) is useful for treating hepatitis caused by HCV. The
CC present sequence represents a peptide given in an example from the
CC present invention
XX
SQ Sequence 8 AA;

Query Match 97.1%; Score 34; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
|||||||
Db 1 DTEDVVA 7

RESULT 9
AAE10047
ID AAE10047 standard; peptide; 10 AA.
XX
AC AAE10047;
XX
DT 29-NOV-2001 (first entry)
XX
DE Hepatitis C virus (HCV) nitroanilide based chromogenic substrate #4.
XX
KW Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-acetyl Gly"
FT Modified-site 8
FT /note= "Cys modified with 2-chloro-4-nitroanilide"
XX
PN US6251583-B1.
XX
PD 26-JUN-2001.
XX
PF 08-APR-1999; 99US-00288391.
XX
PR 27-APR-1998; 98US-0083204P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Malcolm BA, Beyer BM, Njoroge FG, Durkin JP, Windsor WT;
XX
DR WPI; 2001-556521/62.
XX
PT New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS Claim 8; Col 17; 21pp; English.
XX
CC The invention relates to a chromogenic, fluorogenic and fluorescence
CC polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC single chromophore or fluorophore linked to the C-terminus of a peptide
CC sequence, or a fluorescence polarisation HCV substrate comprising a
CC peptide sequence linked at opposite ends of the cleavage site to a
CC fluorophore and a high molecular weight binding group. The chromogenic,
CC fluorogenic and fluorescence polarisation peptide substrates provide
CC optimised specificity, better cleavage efficiency and improved
CC detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC peptide substrates are useful in discovering inhibitors of HCV proteases,
CC in progress curve analysis for reversible and irreversible binding
CC inhibitors for the HCV NS3 protease. These substrates may also be used in
CC monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC protease inhibitors, and to aid in the classification of inhibitors
CC binding to either the S or S' pocket. The present sequence is HCV
CC nitroanilide based chromogenic substrate
XX
SQ Sequence 10 AA;

Query Match 97.1%; Score 34; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
|||||||
Db 3 DTEDVVA 9

RESULT 10
AAE10058
ID AAE10058 standard; peptide; 10 AA.
XX
AC AAE10058;
XX
DT 29-NOV-2001 (first entry)

```
XX
DE   Hepatitis C virus (HCV) fluorogenic substrate #2.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   10
FT                   /label= Nva
FT                   /note= "Nva modified with 9-hydroxy-4-methoxyacridine
FT                   ester"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 19; Col 9; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   fluorogenic substrate
XX
SQ   Sequence 10 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          3 DTEDVVA 9


RESULT 11
AAE10052
ID   AAE10052 standard; peptide; 10 AA.
XX
AC   AAE10052;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #9.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   8
FT                   /label= Nva
FT                   /note= "Nva-O-3-nitrophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 10 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          3 DTEDVVA 9


RESULT 12
AAE10050
ID   AAE10050 standard; peptide; 12 AA.
XX
AC   AAE10050;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Hepatitis C virus nitrophenol and ester based chromogenic substrate #7.
XX
KW   Hepatitis C virus; HCV protease; chromogenic substrate; fluorophore;
KW   chromophore; fluorogenic; fluorescence polarisation substrate.
XX
OS   Hepatitis C virus.
OS   Synthetic.
XX
```

```
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-acetyl Asp"
FT   Modified-site   12
FT                   /label= Nva
FT                   /note= "Nva-O-4-nitrophenol"
XX
PN   US6251583-B1.
XX
PD   26-JUN-2001.
XX
PF   08-APR-1999;   99US-00288391.
XX
PR   27-APR-1998;   98US-0083204P.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Malcolm BA,  Beyer BM,  Njoroge FG,  Durkin JP,  Windsor WT;
XX
DR   WPI; 2001-556521/62.
XX
PT   New chromogenic, fluorogenic and fluorescence polarization hepatitis C
PT   virus (HCV) substrates useful in HCV NS3 protease and inhibitor assays.
XX
PS   Claim 8; Col 17; 21pp; English.
XX
CC   The invention relates to a chromogenic, fluorogenic and fluorescence
CC   polarisation hepatitis C virus (HCV) substrate. The substrate comprises a
CC   single chromophore or fluorophore linked to the C-terminus of a peptide
CC   sequence, or a fluorescence polarisation HCV substrate comprising a
CC   peptide sequence linked at opposite ends of the cleavage site to a
CC   fluorophore and a high molecular weight binding group. The chromogenic,
CC   fluorogenic and fluorescence polarisation peptide substrates provide
CC   optimised specificity, better cleavage efficiency and improved
CC   detectability. The chromogenic, fluorogenic and fluorescence polarisation
CC   peptide substrates are useful in discovering inhibitors of HCV proteases,
CC   in progress curve analysis for reversible and irreversible binding
CC   inhibitors for the HCV NS3 protease. These substrates may also be used in
CC   monitoring of inhibition kinetics and rapid characterisation of HCV NS3
CC   protease inhibitors, and to aid in the classification of inhibitors
CC   binding to either the S or S' pocket. The present sequence is HCV
CC   nitroanilide based chromogenic substrate
XX
SQ   Sequence 12 AA;

  Query Match             97.1%;  Score 34;  DB 4;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          5 DTEDVVA 11


RESULT 13
AAW12958
ID   AAW12958 standard; peptide; 16 AA.
XX
AC   AAW12958;
XX
DT   29-MAR-1997  (first entry)
XX
DE   HCV mutant 5A/5B substrate.
XX
KW   Hepatitis C virus; HCV; NS3 protease; substrate;
KW   nonstructural polyprotein; inhibitor; assay; liver disease;
KW   hepatocellular carcinoma; tumour.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 17
FT                   /note= "residue 7 is Cys in the native 5A/5B substrate"
XX
PN   WO9635717-A2.
XX
PD   14-NOV-1996.
XX
PF   09-MAY-1996;   96WO-US006389.
XX
PR   12-MAY-1995;   95US-00439747.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Zhang R,  Murray MG,  Ramanathan L;
XX
DR   WPI; 1996-518617/51.
XX
PT   New soluble substrates for hepatitis C virus NS3 protease - are non-
PT   structural poly:proteins and are attached to solubilising motifs, useful
PT   for determining protease inhibitors.
XX
PS   Claim 5; Page 57; 70pp; English.
XX
CC   Substrate peptides (AAW12957-62) for hepatitis C virus (HCV) protease NS3
CC   comprise nonstructural polyprotein cleavage sites (NS5A/5B or 4B/5A) of
CC   HCV. The native substrates (see also AAW09251) are made soluble by
CC   attachment of solubilising motifs. The peptides can be synthesised in
CC   large quantities for use in high throughput screens (e.g. scintillation
CC   proximity assay or surface plasmon resonance) to identify protease
CC   inhibitors of therapeutic appln., and for structural studies. NS3
CC   protease (AAW12963) may also be produced in modified form (see also
CC   AAW12964-66 and AAW09236-41) for use in the screening assays
XX
SQ   Sequence 16 AA;

  Query Match             97.1%;  Score 34;  DB 2;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 2.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 14
AAW01653
ID   AAW01653 standard; peptide; 16 AA.
XX
AC   AAW01653;
XX
DT   21-APR-1997  (first entry)
XX
DE   Mutant 5A/5B substrate.
XX
KW   HCV; NS3 protease; inhibitor; scintillation proximity assay.
XX
OS   Synthetic.
XX
PN   WO9636702-A2.
XX
PD   21-NOV-1996.
XX
PF   09-MAY-1996;   96WO-US006387.
XX
PR   12-MAY-1995;   95US-00440409.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Dasmahapatra B,  Murray MG,  Ramanathan L,  Butkiewicz NJ;
XX
DR   WPI; 1997-012081/01.
XX
PT   Bacterially produced Hepatitis C virus NS3 protease(s) - which are
PT   denatured and re-folded to produce soluble, active enzyme.
XX
```

PS Example 3; Page 58; 71pp; English.
XX
CC Hepatitis C virus (HCV) 4B/5A and 5A/5B synthetic substrates (AAW01652-
CC 56, AAW00875) are utilised in high throughput (e.g. scintillation
CC proximity) assays to screen for cpds. that inhibit HCV NS3 protease
CC activity. The virus cannot replicate if the ability of the NS3 protease
CC to cleave the substrate is inhibited. Soluble, active NS3 proteases (see
CC also AAW01643-50) are preferably used in the assay. Protease inhibitors
CC may have therapeutic value against HCV
XX
SQ Sequence 16 AA;

Query Match 97.1%; Score 34; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
||||||| |
Db 1 DTEDVVA 7

RESULT 15
AAW12959
ID AAW12959 standard; peptide; 17 AA.
XX
AC AAW12959;
XX
DT 29-MAR-1997 (first entry)
XX
DE HCV mutant soluble 5A/5B substrate.
XX
KW Hepatitis C virus; HCV; NS3 protease; substrate;
KW nonstructural polyprotein; inhibitor; assay; liver disease;
KW hepatocellular carcinoma; tumour.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 7
FT /note= "residue 7 is Cys in the native 5A/5B substrate"
FT Misc-difference 17
FT /note= "C-terminal Lys acts as a solubilising motif"
XX
PN WO9635717-A2.
XX
PD 14-NOV-1996.
XX
PF 09-MAY-1996; 96WO-US006389.
XX
PR 12-MAY-1995; 95US-00439747.
XX
PA (SCHE) SCHERING CORP.
XX
PI Zhang R, Murray MG, Ramanathan L;
XX
DR WPI; 1996-518617/51.
XX
PT New soluble substrates for hepatitis C virus NS3 protease - are non-
PT structural poly:proteins and are attached to solubilising motifs, useful
PT for determining protease inhibitors.
XX
PS Claim 5; Page 58; 70pp; English.
XX
CC Substrate peptides (AAW12957-62) for hepatitis C virus (HCV) protease NS3
CC comprise nonstructural polyprotein cleavage sites (NS5A/5B or 4B/5A) of
CC HCV. The native substrates (see also AAW09251) are made soluble by
CC attachment of solubilising motifs. The peptides can be synthesised in
CC large quantities for use in high throughput screens (e.g. scintillation
CC proximity assay or surface plasmon resonance) to identify protease
CC inhibitors of therapeutic appln., and for structural studies. NS3
CC protease (AAW12963) may also be produced in modified form (see also
CC AAW12964-66 and AAW09236-41) for use in the screening assays
XX
SQ Sequence 17 AA;

Query Match 97.1%; Score 34; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
||||||| |
Db 1 DTEDVVA 7

Search completed: March 31, 2004, 16:45:33
Job time : 50.6667 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:42:17 ; Search time 14.1333 Seconds
(without alignments)
29.222 Million cell updates/sec

Title: US-09-909-077-3
Perfect score: 35
Sequence: 1 DTEDVVAX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	97.1	8	3	US-09-288-391-3	Sequence 3, Appli
2	34	97.1	8	3	US-09-288-391-5	Sequence 5, Appli
3	34	97.1	8	3	US-09-288-391-6	Sequence 6, Appli
4	34	97.1	8	3	US-09-288-391-8	Sequence 8, Appli
5	34	97.1	8	3	US-09-288-391-10	Sequence 10, Appl
6	34	97.1	8	3	US-09-288-391-12	Sequence 12, Appl
7	34	97.1	10	3	US-09-288-391-4	Sequence 4, Appli
8	34	97.1	10	3	US-09-288-391-9	Sequence 9, Appli
9	34	97.1	10	3	US-09-288-391-15	Sequence 15, Appl
10	34	97.1	12	3	US-09-288-391-7	Sequence 7, Appli
11	34	97.1	16	1	US-08-439-747A-17	Sequence 17, Appl
12	34	97.1	16	2	US-08-440-409B-17	Sequence 17, Appl
13	34	97.1	17	1	US-08-571-643A-5	Sequence 5, Appli
14	34	97.1	17	1	US-08-439-747A-18	Sequence 18, Appl
15	34	97.1	17	2	US-08-440-409B-18	Sequence 18, Appl
16	34	97.1	17	3	US-09-198-723A-91	Sequence 91, Appl
17	34	97.1	17	3	US-09-288-391-24	Sequence 24, Appl
18	34	97.1	17	4	US-09-684-881-91	Sequence 91, Appl
19	34	97.1	18	1	US-08-439-747A-20	Sequence 20, Appl
20	34	97.1	18	2	US-08-440-409B-20	Sequence 20, Appl
21	30	85.7	7	4	US-09-777-785A-1	Sequence 1, Appli
22	30	85.7	8	3	US-09-288-391-11	Sequence 11, Appl
23	30	85.7	8	3	US-09-288-391-13	Sequence 13, Appl
24	30	85.7	8	3	US-09-288-391-14	Sequence 14, Appl
25	30	85.7	10	3	US-09-288-391-1	Sequence 1, Appli
26	30	85.7	12	3	US-09-288-391-2	Sequence 2, Appli
27	30	85.7	13	3	US-09-288-391-16	Sequence 16, Appl
28	30	85.7	13	3	US-09-288-391-18	Sequence 18, Appl
29	30	85.7	13	3	US-09-288-391-19	Sequence 19, Appl
30	30	85.7	14	3	US-09-288-391-17	Sequence 17, Appl
31	30	85.7	14	4	US-09-344-456-3	Sequence 3, Appli
32	30	85.7	16	1	US-08-439-747A-31	Sequence 31, Appl
33	30	85.7	16	2	US-08-853-623D-25	Sequence 25, Appl
34	30	85.7	16	3	US-09-198-723A-25	Sequence 25, Appl
35	30	85.7	16	4	US-09-684-881-25	Sequence 25, Appl
36	30	85.7	17	1	US-08-439-747A-16	Sequence 16, Appl
37	30	85.7	17	2	US-08-440-409B-16	Sequence 16, Appl
38	30	85.7	17	2	US-08-853-623D-22	Sequence 22, Appl
39	30	85.7	17	3	US-09-288-391-20	Sequence 20, Appl
40	30	85.7	17	3	US-09-288-391-21	Sequence 21, Appl
41	30	85.7	18	1	US-08-439-747A-19	Sequence 19, Appl
42	30	85.7	18	2	US-08-440-409B-19	Sequence 19, Appl
43	30	85.7	18	3	US-09-144-759-10	Sequence 10, Appl
44	30	85.7	18	4	US-09-570-267-10	Sequence 10, Appl
45	30	85.7	20	2	US-08-432-693-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-09-288-391-3
; Sequence 3, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583e1 Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The aspartic acid residue at position 1 is N-acety
US-09-288-391-3

  Query Match             97.1%;  Score 34;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 2
US-09-288-391-5
; Sequence 5, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Abu (a-aminobutyric acid).  T
US-09-288-391-5

  Query Match             97.1%;  Score 34;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 3
US-09-288-391-6
; Sequence 6, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-6

  Query Match             97.1%;  Score 34;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 4
US-09-288-391-8
; Sequence 8, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
```

```
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McLaughlin, Jaye P.
;       REGISTRATION NUMBER:  41,211
;       REFERENCE/DOCKET NUMBER:  IN0829P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908)298-5056
;       TELEFAX:  (908)298-5388
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-8

  Query Match             97.1%;  Score 34;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 5
US-09-288-391-10
; Sequence 10, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline).  The asparti
US-09-288-391-10

  Query Match             97.1%;  Score 34;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 6
US-09-288-391-12
; Sequence 12, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 8 is Nva (norvaline). The aspartic
US-09-288-391-12

  Query Match             97.1%;  Score 34;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 7
US-09-288-391-4
; Sequence 4, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
```

```
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= The glycine residue at position 1 is N-acetylated.
US-09-288-391-4

  Query Match             97.1%;  Score 34;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          3 DTEDVVA 9


RESULT 8
US-09-288-391-9
; Sequence 9, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 10 is Nva (norvaline).  The Nva re
US-09-288-391-9

  Query Match             97.1%;  Score 34;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          3 DTEDVVA 9


RESULT 9
US-09-288-391-15
; Sequence 15, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  /note= Xaa at position 10 is Nva (norvaline). The Nva res
US-09-288-391-15

  Query Match             97.1%;  Score 34;  DB 3;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          3 DTEDVVA 9


RESULT 10
US-09-288-391-7
; Sequence 7, Application US/09288391
; Patent No. 6251583
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Malcolm, Bruce
;    APPLICANT:  Beyer, Brian
;    APPLICANT:  Njoroge, George
;    APPLICANT:  Durkin, James
;    APPLICANT:  Windsor, William
;    TITLE OF INVENTION:  No. 6251583el Peptide Substrates for HCV NS3 Protease Assay
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/288,391
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLaughlin, Jaye P.
;      REGISTRATION NUMBER:  41,211
;      REFERENCE/DOCKET NUMBER:  IN0829P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)298-5056
;      TELEFAX:  (908)298-5388
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  /note= Xaa at position 12 is Nva (norvaline).  The serine
US-09-288-391-7

  Query Match             97.1%;  Score 34;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          5 DTEDVVA 11


RESULT 11
US-08-439-747A-17
; Sequence 17, Application US/08439747A
; Patent No. 5767233
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Murray, Michael
;    APPLICANT:  Ramanathan, Lata
;    TITLE OF INVENTION:  Soluble, Cleavable Substrates of the Hepatitis
;    TITLE OF INVENTION:  C Protease
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033-0530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5.3
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,747A
;      FILING DATE:  May 12, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Paul G.
;      REGISTRATION NUMBER:  32,743
;      REFERENCE/DOCKET NUMBER:  JB0509
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-298-5061
;      TELEFAX:  908-298-5388
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  polypeptide
;    FEATURE:
;      NAME/KEY:  Mutant 5A/5B Substrate
US-08-439-747A-17

  Query Match             97.1%;  Score 34;  DB 1;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 12
US-08-440-409B-17
; Sequence 17, Application US/08440409B
; Patent No. 5843752
;  GENERAL INFORMATION:
;    APPLICANT:  Dasmahapatra, Bimal
;    APPLICANT:  Murray, Michael
;    APPLICANT:  Ramanathan, Lata
;    APPLICANT:  Butkiewicz, Nancy
;    TITLE OF INVENTION:  Soluble Active Hepatitis C Virus Protease
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033-0530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  Macintosh 7.5.3
;       SOFTWARE:  Microsoft Word 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/440,409B
;       FILING DATE:  May 12, 1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lunn, Paul G.
;       REGISTRATION NUMBER:  32,743
;       REFERENCE/DOCKET NUMBER:  JB0494
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  908-298-5061
;       TELEFAX:  908-298-5388
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  16 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  polypeptide
;     FEATURE:
;       NAME/KEY:  Mutant 5A/5B Substrate
US-08-440-409B-17

  Query Match             97.1%;  Score 34;  DB 2;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 13
US-08-571-643A-5
; Sequence 5, Application US/08571643A
; Patent No. 5714371
;  GENERAL INFORMATION:
;    APPLICANT:  Ramanthan, Lata
;    APPLICANT:  Wendel, Michelle
;    TITLE OF INVENTION:  Method for Refolding Insoluble
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering-Plough Corporation
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033-0530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.1
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/571,643A
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 08/439,680
;      FILING DATE:  12-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dulak, No. 5714371man C.
;      REGISTRATION NUMBER:  31,308
;      REFERENCE/DOCKET NUMBER:  JB0508K
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-298-2906
;      TELEFAX:  908-298-5388
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  polypeptide
;    FEATURE:
;      NAME/KEY:  Mutant Soluble 5A/5B Substrate
US-08-571-643A-5

  Query Match             97.1%;  Score 34;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.82;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 14
US-08-439-747A-18
; Sequence 18, Application US/08439747A
; Patent No. 5767233
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Rumin
;    APPLICANT:  Murray, Michael
;    APPLICANT:  Ramanathan, Lata
;    TITLE OF INVENTION:  Soluble, Cleavable Substrates of the Hepatitis
;    TITLE OF INVENTION:  C Protease
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  2000 Galloping Hill Road
;      CITY:  Kenilworth
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07033-0530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5.3
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,747A
;      FILING DATE:  May 12, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Paul G.
;      REGISTRATION NUMBER:  32,743
;      REFERENCE/DOCKET NUMBER:  JB0509
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-298-5061
;      TELEFAX:  908-298-5388
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  polypeptide
;    FEATURE:
;      NAME/KEY:  Mutant Soluble 5A/5B Substrate
US-08-439-747A-18

  Query Match             97.1%;  Score 34;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.82;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||||||
Db          1 DTEDVVA 7


RESULT 15
US-08-440-409B-18
; Sequence 18, Application US/08440409B
```

; Patent No. 5843752
; GENERAL INFORMATION:
; APPLICANT: Dasmahapatra, Bimal
; APPLICANT: Murray, Michael
; APPLICANT: Ramanathan, Lata
; APPLICANT: Butkiewicz, Nancy
; TITLE OF INVENTION: Soluble Active Hepatitis C Virus Protease
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5.3
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,409B
; FILING DATE: May 12, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: JB0494
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-5061
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; FEATURE:
; NAME/KEY: Mutant Soluble 5A/5B Substrate
US-08-440-409B-18

Query Match 97.1%; Score 34; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
||||||
Db 1 DTEDVVA 7

Search completed: March 31, 2004, 16:50:33
Job time : 14.1333 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:41:17 ; Search time 11.2 Seconds
(without alignments)
68.708 Million cell updates/sec

Title: US-09-909-077-3
Perfect score: 35
Sequence: 1 DTEDVVAX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	88.6	439	2	G88103	protein W10G11.17
2	30	85.7	371	1	QQVZA7	A7L protein - vacc
3	30	85.7	372	2	D72164	A7L protein - vari
4	30	85.7	372	2	H42517	A6L protein - vacc
5	30	85.7	372	2	G36848	A6L protein - vari
6	30	85.7	372	2	T28548	hypothetical prote
7	30	85.7	372	2	T37393	probable 43.1K pro
8	30	85.7	412	2	T15214	hypothetical prote
9	30	85.7	424	1	A36000	sperm-binding glyc
10	30	85.7	828	2	E71906	DNA gyrase chain A
11	30	85.7	832	2	B87673	ABC transporter, H
12	30	85.7	2135	2	T14602	variant-specific s
13	30	85.7	3011	1	GNWVCH	genome polyprotein
14	29	82.9	200	2	D85574	probable corrinoid
15	29	82.9	200	2	D90723	probable cob(I)ala
16	29	82.9	242	2	H70980	probable rsbW prot
17	29	82.9	244	2	A64820	transaldolase-like
18	29	82.9	244	2	G90741	probable transaldo
19	29	82.9	244	2	B85592	probable transaldo
20	29	82.9	288	2	D85072	hypothetical prote
21	29	82.9	295	2	T24036	hypothetical prote
22	29	82.9	321	2	S55429	glycerol-inducible
23	29	82.9	326	2	C90187	hypothetical prote
24	29	82.9	618	2	A75469	conserved hypothet
25	29	82.9	1049	2	A27079	fibronectin recept
26	29	82.9	1053	2	S44250	integrin alpha-5 c
27	28	80.0	161	2	S67178	translation initia
28	28	80.0	166	2	D69695	ribosomal protein
29	28	80.0	225	2	G84310	cobalamin adenosyl
30	28	80.0	301	2	B71330	hypothetical prote
31	28	80.0	318	2	T41838	BRO-c - Bombyx mor
32	28	80.0	354	2	E72640	hypothetical prote
33	28	80.0	358	1	H64937	probable alcohol d
34	28	80.0	358	2	E90939	probable oxidoredu
35	28	80.0	358	2	A85788	probable oxidoredu
36	28	80.0	377	2	B87537	hypothetical prote
37	28	80.0	400	2	B81285	phosphoglycerate k
38	28	80.0	417	2	B82542	conserved hypothet
39	28	80.0	482	2	A38533	transcription acti
40	28	80.0	487	2	I64033	hypothetical prote
41	28	80.0	508	2	T37224	hypothetical prote
42	28	80.0	526	2	T21811	hypothetical prote
43	28	80.0	530	2	S38092	hypothetical prote
44	28	80.0	559	2	H84586	hypothetical prote
45	28	80.0	583	2	T34121	steroid/thyroid/re

ALIGNMENTS

RESULT 1
G88103
protein W10G11.17 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: G88103
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_el
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
A;Accession: G88103
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-439 <STO>
A;Cross-references: GB:chr_II; PIDN:AB95081.1; PID:g2746939; GSPDB:GN00020; CESP:W10G11
C;Genetics:
A;Gene: W10G11.17
A;Map position: 2

Query Match 88.6%; Score 31; DB 2; Length 439;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVA 7
              |||||| :
Db         40 DTEDVVS 46
```

RESULT 2
QQVZA7
A7L protein - vaccinia virus (strain WR)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 08-Apr-1994
C;Accession: C41806
R;Ahn, B.Y.; Rosel, J.; Cole, N.B.; Moss, B.
J. Virol. 66, 971-982, 1992
A;Title: Identification and expression of rpo19, a vaccinia virus gene encoding a 19-ki
A;Reference number: A41806; MUID:92114202; PMID:1731116
A;Accession: C41806
A;Molecule type: DNA
A;Residues: 1-371 <AHN>
A;Cross-references: GB:M76473
C;Superfamily: vaccinia virus A7L protein

Query Match 85.7%; Score 30; DB 1; Length 371;
Best Local Similarity 71.4%; Pred. No. 71;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVA 7
```

```
              ||||:|:
Db        126 DTEDIVS 132


RESULT 3
D72164
A7L protein - variola minor virus (strain Garcia-1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: D72164
R;Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of alastrim variola minor
A;Reference number: A72150
A;Accession: D72164
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <SHC>
A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54710.1; PID:g5830671
A;Experimental source: strain Garcia-1966
C;Genetics:
A;Gene: A7L
C;Superfamily: vaccinia virus A7L protein

  Query Match             85.7%;  Score 30;  DB 2;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 71;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 4
H42517
A6L protein - vaccinia virus (strain Copenhagen)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 08-Apr-1994
C;Accession: H42517
R;Johnson, G.P.
submitted to GenBank, June 1990
A;Reference number: A33172
A;Accession: H42517
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <JOH>
C;Superfamily: vaccinia virus A7L protein

  Query Match             85.7%;  Score 30;  DB 2;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 71;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 5
G36848
A6L protein - variola virus (strain India-1967)
C;Species: variola virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C;Accession: G36848; S46892
R;Blinov, V.M.
submitted to GenBank, November 1992
A;Reference number: A36859
A;Accession: G36848
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CAA49051.1; PID:g297289
R;Volchov, V.E.; Blinov, V.M.; Totmenin, A.V.; Shchelkunov, S.N.; Sandakhchiev, L.S.
submitted to the EMBL Data Library, April 1992
A;Description: Nucleotide sequence analysis of the region of variola virus XhoI-G genome
A;Reference number: S46890
A;Accession: S46892
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <VOL>
A;Cross-references: EMBL:X67116; NID:g516451; PIDN:CAA47514.1; PID:g516454
C;Superfamily: vaccinia virus A7L protein

  Query Match             85.7%;  Score 30;  DB 2;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 71;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 6
T28548
hypothetical protein A7L - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28548
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of variola smallpox virus
A;Reference number: Z20488; MUID:94088747; PMID:8264798
A;Accession: T28548
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-372 <MAS>
A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60858.1; PID:g439028
A;Experimental source: strain Bangladesh-1975
C;Superfamily: vaccinia virus A7L protein

  Query Match             85.7%;  Score 30;  DB 2;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 71;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 7
T37393
probable 43.1K protein - vaccinia virus (strain Ankara)
C;Species: vaccinia virus
A;Variety: strain Ankara
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C;Accession: T37393
R;Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.
submitted to the EMBL Data Library, March 1997
A;Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strai
A;Reference number: Z20877
A;Accession: T37393
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-372 <ANT>
A;Cross-references: EMBL:U94848; PIDN:AAB96459.1
A;Experimental source: strain Ankara
C;Genetics:
A;Note: MVA117L
C;Superfamily: vaccinia virus A7L protein

  Query Match             85.7%;  Score 30;  DB 2;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 71;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 16:37:40 ; Search time 7.46667 Seconds
(without alignments)
55.789 Million cell updates/sec

Title: US-09-909-077-3
Perfect score: 35
Sequence: 1 DTEDVVAX 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	85.7	371	1	VA06_VACCV	P29192 vaccinia vi
2	30	85.7	372	1	VA06_VACCC	P20985 vaccinia vi
3	30	85.7	372	1	VA06_VARV	P33833 variola vir
4	30	85.7	424	1	ZP3_CALSQ	P53786 callithrix
5	30	85.7	424	1	ZP3_HUMAN	P21754 homo sapien
6	30	85.7	424	1	ZP3_MACRA	P53785 macaca radi
7	30	85.7	828	1	GYRA_HELPJ	Q9zld9 helicobacte
8	30	85.7	3011	1	POLG_HCVH	P27958 h genome po
9	29	82.9	220	1	FSAA_ECO57	P58423 escherichia
10	29	82.9	220	1	FSAA_ECOLI	P78055 escherichia
11	29	82.9	321	1	YWJI_BACSU	Q03224 bacillus su
12	29	82.9	459	1	TRME_STAEP	Q8cmn5 staphylococ
13	29	82.9	491	1	2A5R_MOUSE	Q9z176 mus musculu
14	29	82.9	1049	1	ITA5_HUMAN	P08648 homo sapien
15	29	82.9	1053	1	ITA5_MOUSE	P11688 mus musculu
16	28	80.0	165	1	RL10_BACSU	P42923 bacillus su
17	28	80.0	166	1	RL10_STAEP	Q8ctt2 staphylococ
18	28	80.0	301	1	Y376_TREPA	O83391 treponema p
19	28	80.0	324	1	MURB_RHIME	Q92nm1 rhizobium m
20	28	80.0	358	1	YDJL_ECOLI	P77539 escherichia
21	28	80.0	400	1	PGK_CAMJE	Q9pmq5 campylobact
22	28	80.0	417	1	ISPG_XYLFA	Q9pae3 xylella fas
23	28	80.0	417	1	ISPG_XYLFT	Q87a73 xylella fas
24	28	80.0	482	1	HOXA_ALCEU	P29267 alcaligenes
25	28	80.0	487	1	VPL_HAEIN	P44233 haemophilus
26	28	80.0	530	1	YK03_YEAST	P36119 saccharomyc
27	28	80.0	583	1	NH31_CAEEL	Q18192 caenorhabdi
28	28	80.0	589	1	SYD_MYCLE	P36429 mycobacteri
29	28	80.0	653	1	YD64_MYCTU	Q11034 mycobacteri
30	28	80.0	736	1	ALD_MOUSE	P48410 mus musculu
31	28	80.0	745	1	ALD_HUMAN	P33897 homo sapien
32	28	80.0	856	1	UN51_CAEEL	Q23023 caenorhabdi
33	28	80.0	899	1	R24L_ARATH	Q9c646 arabidopsis
34	28	80.0	2376	1	TAO3_YEAST	P40468 saccharomyc
35	28	80.0	5120	1	PCLO_CHICK	Q9pu36 gallus gall
36	27	77.1	105	1	FLHD_RALSO	Q8xq73 ralstonia s
37	27	77.1	113	1	MEA1_PIG	Q95313 sus scrofa
38	27	77.1	174	1	MEA1_BOVIN	Q29407 bos taurus
39	27	77.1	181	1	KAD_MYCLE	O33007 mycobacteri
40	27	77.1	203	1	KGUA_XANCP	Q8p5t7 xanthomonas
41	27	77.1	286	1	3HAO_RAT	P46953 rattus norv
42	27	77.1	304	1	RESA_PLAFP	Q26005 plasmodium
43	27	77.1	319	1	YQJP_BACSU	P54553 bacillus su
44	27	77.1	321	1	MURB_AGRT5	Q8udn0 agrobacteri
45	27	77.1	335	1	KG62_MOUSE	Q8c010 mus musculu

ALIGNMENTS

RESULT 1
VA06_VACCV
ID VA06_VACCV STANDARD; PRT; 371 AA.
AC P29192;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein A6 (A7).
GN A6L OR A7L.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114202; PubMed=1731116;
RA Ahn B.-Y., Rosel J., Cole N.B., Moss B.;
RT "Identification and expression of rpo19, a vaccinia virus gene
RT encoding a 19-kilodalton DNA-dependent RNA polymerase subunit.";
RL J. Virol. 66:971-982(1992).
CC -!- SIMILARITY: BELONGS TO THE POXVIRUSES A6 FAMILY.
CC ---

CC ---
DR EMBL; M76473; -; NOT_ANNOTATED_CDS.
DR PIR; C41806; QQVZA7.
DR InterPro; IPR007008; Pox_A6.
DR Pfam; PF04924; Pox_A6; 1.
SQ SEQUENCE 371 AA; 43006 MW; 013A21BD94612AC6 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 371;
Best Local Similarity 71.4%; Pred. No. 41;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DTEDVVA 7
              ||||:|:
Db        126 DTEDIVS 132
```

RESULT 2
VA06_VACCC
ID VA06_VACCC STANDARD; PRT; 372 AA.
AC P20985;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein A6.
GN A6L.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

```
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35027; AAA48123.1; -.
DR   PIR; H42517; H42517.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43127 MW;  8149CD07AD808D70 CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 41;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 3
VA06_VARV
ID   VA06_VARV      STANDARD;      PRT;   372 AA.
AC   P33833;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A6.
GN   A6L OR A7L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=India-1967 / Isolate Ind3;
RX   MEDLINE=92209372; PubMed=1666548;
RA   Shchelkunov S.N., Marennikova S.S., Totmenin A.V., Blinov V.M.,
RA   Chizhikov V.E., Gutorov V.V., Safronov P.F., Pozdnyakov S.G.,
RA   Shelukhina E.M., Gashnikov P.V., Anjaparidze O.G., Sandakhchiev L.S.;
RT   "Creation of a clone library of fragments from the natural variola
RT   virus and study of the structural and functional organization of
RT   viral genes from a circle of hosts.";
RL   Dokl. Akad. Nauk SSSR 321:402-406(1991).
RN   [2]
RP   COMPLETE GENOME.
RC   STRAIN=India-1967 / Isolate Ind3;
RX   MEDLINE=93202281; PubMed=8384129;
RA   Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT   "Genes of variola and vaccinia viruses necessary to overcome the host
RT   protective mechanisms.";
RL   FEBS Lett. 319:80-83(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bangladesh-1975;
RX   MEDLINE=94088747; PubMed=8264798;
RA   Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA   Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA   Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA   Venter C.J.;
RT   "Potential virulence determinants in terminal regions of variola
RT   smallpox virus genome.";
RL   Nature 366:748-751(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Garcia-1966;
RA   Shchelkunov S.N., Totmenin A.V., Safronov P.F., Resenchuk S.M.,
RA   Blinov V.M., Sandakhchiev L.S.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69198; CAA49051.1; -.
DR   EMBL; X67116; CAA47514.1; -.
DR   EMBL; L22579; AAA60858.1; -.
DR   EMBL; X76265; CAA53848.1; -.
DR   PIR; D72164; D72164.
DR   PIR; G36848; G36848.
DR   PIR; T28548; T28548.
DR   InterPro; IPR007008; Pox_A6.
DR   Pfam; PF04924; Pox_A6; 1.
SQ   SEQUENCE   372 AA;  43148 MW;  9751173CE363452D CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 372;
  Best Local Similarity   71.4%;  Pred. No. 41;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||:|:
Db        127 DTEDIVS 133


RESULT 4
ZP3_CALSQ
ID   ZP3_CALSQ      STANDARD;      PRT;   424 AA.
AC   P53786;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor).
GN   ZP3.
OS   Callithrix sp. (Marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC   Callithrix.
OX   NCBI_TaxID=9485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=94363314; PubMed=8081814;
RA   Thillai-Koothan P., van Duin M., Aitken R.J.;
RT   "Cloning, sequencing and oocyte-specific expression of the marmoset
RT   sperm receptor protein, ZP3.";
RL   Zygote 1:93-101(1993).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
```

```
Db        127 DTEDIVS 133


RESULT 8
T15214
hypothetical protein F57C9.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15214
R;Geisel, C.; Kramer, J.; Gibson, A.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid F57C9.
A;Reference number: Z18309
A;Accession: T15214
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-412 <GEI>
A;Cross-references: EMBL:AF003142; NID:g2088743; PID:g2088750; PIDN:AAB54190.1; GSPDB:GN
A;Experimental source: strain Bristol N2; clone F57C9
C;Genetics:
A;Gene: CESP:F57C9.7
A;Map position: 1
A;Introns: 100/3; 158/3; 194/3; 272/2; 311/2

  Query Match             85.7%;  Score 30;  DB 2;  Length 412;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db        132 DTEDVV 137


RESULT 9
A36000
sperm-binding glycoprotein ZP3 precursor - human
N;Alternate names: sperm receptor ZP3; zona pellucida glycoprotein ZP3
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36000; A44365
R;Chamberlin, M.E.; Dean, J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6014-6018, 1990
A;Title: Human homolog of the mouse sperm receptor.
A;Reference number: A36000; MUID:90349545; PMID:2385582
A;Accession: A36000
A;Molecule type: mRNA; DNA
A;Residues: 1-424 <CHA>
A;Cross-references: GB:M60504; GB:M35109; NID:g340491; PIDN:AAA61336.1; PID:g340492
R;van Duin, M.; Polman, J.E.; Verkoelen, C.C.; Bunschoten, H.; Meyerink, J.H.; Olijve, W
Genomics 14, 1064-1070, 1992
A;Title: Cloning and characterization of the human sperm receptor ligand ZP3: evidence f
A;Reference number: A44365; MUID:93122771; PMID:1478648
A;Accession: A44365
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 329-370,'S',372-424 <VAN>
A;Experimental source: ovary
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:122391, NCBIP:122392)
C;Comment: This sulfated glycoprotein in the zona pellucida of the oocyte is a receptor
C;Genetics:
A;Gene: GDB:ZP3A
A;Cross-references: GDB:128007; OMIM:182889
A;Map position: 7pter-7qter
C;Superfamily: sperm-binding glycoprotein ZP3; ZP domain homology
C;Keywords: glycoprotein; oocyte; receptor; sulfoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-424/Product: sperm-binding glycoprotein ZP3 #status predicted <MAT>
F;45-301/Domain: ZP domain homology <ZPH>

  Query Match             85.7%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 10
E71906
DNA gyrase chain A - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 21-Jan-2000
C;Accession: E71906
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: E71906
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-828 <ARN>
A;Cross-references: GB:AE001496; GB:AE001439; NID:g4155191; PIDN:AAD06219.1; PID:g41551
A;Experimental source: strain J99
C;Genetics:
A;Gene: gyrA
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase
F;5-242/Domain: phage T4 DNA topoisomerase (ATP-hydrolyzing) medium chain homology <T4T

  Query Match             85.7%;  Score 30;  DB 2;  Length 828;
  Best Local Similarity   71.4%;  Pred. No. 1.8e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||::|
Db        497 DTEDLIA 503


RESULT 11
B87673
ABC transporter, HlyB/MsbA family CC3420 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: B87673
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87673
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-832 <STO>
A;Cross-references: GB:AE005673; NID:g13425134; PIDN:AAK25382.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC3420

  Query Match             85.7%;  Score 30;  DB 2;  Length 832;
  Best Local Similarity   85.7%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||| |
Db        562 DTEDVAA 568


RESULT 12
T14602
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C;Accession: T14602
```

R;Voss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998
A;Description: Identification of a conserved 5' flanking region of Plasmodium falciparum
A;Reference number: Z18158
A;Accession: T14602
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2135 <VOS>
A;Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1
C;Genetics:
A;Gene: varph17

Query Match 85.7%; Score 30; DB 2; Length 2135;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 1324 DTEDVV 1329

RESULT 13
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: A36814; A41546
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 85.7%; Score 30; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVV 6
||||||
Db 2413 DTEDVV 2418

RESULT 14
D85574
probable corrinoid,ATP adenosyltransferase Z0886 [imported] - Escherichia coli (strain O
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 28-Jul-2003
C;Accession: D85574
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-200 <STO>
A;Cross-references: GB:AE005174; NID:g12513650; PIDN:AAG55056.1; GSPDB:GN00145; UWGP:Z08
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0886
C;Superfamily: ATP:cob(I)alamin adenosyltransferase, CobA type

Query Match 82.9%; Score 29; DB 2; Length 200;
Best Local Similarity 71.4%; Pred. No. 59;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
|||:|:|
Db 143 DTEEVIA 149

RESULT 15
D90723
probable cob(I)alamin adenosyltransferase [imported] - Escherichia coli (strain O157:H7,
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 28-Jul-2003
C;Accession: D90723
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90723
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-200 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34179.1; PID:g13360215; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0756
C;Superfamily: ATP:cob(I)alamin adenosyltransferase, CobA type

Query Match 82.9%; Score 29; DB 2; Length 200;
Best Local Similarity 71.4%; Pred. No. 59;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTEDVVA 7
|||:|:|
Db 143 DTEEVIA 149

Search completed: March 31, 2004, 16:49:32
Job time : 13.2 secs

```
                |||| :|
Db          83 DTEDAIA 89


RESULT 14
ITA5_HUMAN
ID   ITA5_HUMAN     STANDARD;      PRT;  1049 AA.
AC   P08648; Q96HA5;
DT   01-AUG-1988 (Rel. 08, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Integrin alpha-5 precursor (Fibronectin receptor alpha subunit)
DE   (Integrin alpha-F) (VLA-5) (CD49e).
GN   ITGA5 OR FNRA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88007843; PubMed=2958481;
RA   Argraves W.S., Suzuki S., Arai H., Thompson K., Pierschbacher M.D.,
RA   Ruoslahti E.;
RT   "Amino acid sequence of the human fibronectin receptor.";
RL   J. Cell Biol. 105:1183-1190(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 16-1049 FROM N.A.
RX   MEDLINE=88163472; PubMed=2450560;
RA   Fitzgerald L.A., Poncz M., Steiner B., Rall S.C., Bennett J.S.,
RA   Phillips D.R.;
RT   "Comparison of cDNA-derived protein sequences of the human
RT   fibronectin and vitronectin receptor alpha-subunits and platelet
RT   glycoprotein IIb.";
RL   Biochemistry 26:8158-8165(1987).
RN   [4]
RP   SEQUENCE OF 821-1049 FROM N.A.
RX   MEDLINE=87008489; PubMed=2944883;
RA   Argraves W.S., Pytela R., Suzuki S., Millan J.L., Pierschbacher M.D.,
RA   Ruoslahti E.;
RT   "cDNA sequences from the alpha subunit of the fibronectin receptor
RT   predict a transmembrane domain and a short cytoplasmic peptide.";
RL   J. Biol. Chem. 261:12922-12924(1986).
RN   [5]
RP   SEQUENCE OF 42-55.
RX   MEDLINE=87204112; PubMed=3033641;
RA   Takada Y., Strominger J.L., Hemler M.E.;
RT   "The very late antigen family of heterodimers is part of a
RT   superfamily of molecules involved in adhesion and embryogenesis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:3239-3243(1987).
```

```
CC   -!- FUNCTION: INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRONECTIN
CC       AND FIBRINOGEN. IT RECOGNIZES THE SEQUENCE R-G-D IN ITS LIGANDS.
CC   -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. The alpha
CC       subunit is composed of an heavy and a light chain linked by a
CC       disulfide bond. Alpha-5 associates with beta-1. Interacts with
CC       HPS5.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC   -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD49e entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49e.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06256; CAA29601.1; -.
DR   EMBL; BC008786; AAH08786.1; -.
DR   EMBL; M13918; AAA52467.1; ALT_SEQ.
DR   PIR; A27079; A27079.
DR   HSSP; P06756; 1JV2.
DR   GlycoSuiteDB; P08648; -.
DR   Genew; HGNC:6141; ITGA5.
DR   MIM; 135620; -.
DR   GO; GO:0008305; C:integrin complex; TAS.
DR   GO; GO:0004895; F:cell adhesion receptor activity; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 4.
DR   Pfam; PF00357; integrin_A; 1.
DR   SMART; SM00191; Int_alpha; 5.
DR   PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW   Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW   Signal; Calcium; Repeat.
FT   SIGNAL       1     41
FT   CHAIN       42   1049       INTEGRIN ALPHA-5.
FT   CHAIN       42    894       INTEGRIN ALPHA-5 HEAVY CHAIN.
FT   CHAIN      895   1049       INTEGRIN ALPHA-5 LIGHT CHAIN.
FT   DOMAIN      42    995       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   996   1021       POTENTIAL.
FT   DOMAIN    1022   1049       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN    1021   1028       INTERACTION WITH HPS5.
FT   REPEAT      57    119       FG-GAP 1.
FT   REPEAT     131    198       FG-GAP 2.
FT   REPEAT     199    256       FG-GAP 3.
FT   REPEAT     269    322       FG-GAP 4.
FT   REPEAT     323    389       FG-GAP 5.
FT   REPEAT     390    449       FG-GAP 6.
FT   REPEAT     453    505       FG-GAP 7.
FT   CA_BIND    334    342       POTENTIAL.
FT   CA_BIND    401    409       POTENTIAL.
FT   CA_BIND    465    473       POTENTIAL.
FT   SITE      1024   1028       GFFKR MOTIF.
FT   DISULFID    99    108       BY SIMILARITY.
FT   DISULFID   156    176       BY SIMILARITY.
FT   DISULFID   192    205       BY SIMILARITY.
FT   DISULFID   513    522       BY SIMILARITY.
FT   DISULFID   528    584       BY SIMILARITY.
FT   DISULFID   645    651       BY SIMILARITY.
FT   DISULFID   718    731       BY SIMILARITY.
FT   DISULFID   869    911       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   916    921       BY SIMILARITY.
FT   CARBOHYD    84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   307    307       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   524    524       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    593    593       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    609    609       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    675    675       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    712    712       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    724    724       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    773    773       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    868    868       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     26     26       L -> V (IN REF. 1).
FT   CONFLICT     33     33       L -> V (IN REF. 1).
SQ   SEQUENCE   1049 AA;  114536 MW;  6B4D558D4F739CBA CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 1049;
  Best Local Similarity   85.7%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||| ||
Db        285 DTEDFVA 291


RESULT 15
ITA5_MOUSE
ID   ITA5_MOUSE     STANDARD;      PRT;  1053 AA.
AC   P11688;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Integrin alpha-5 precursor (Fibronectin receptor alpha subunit)
DE   (Integrin alpha-F) (VLA-5) (CD49e).
GN   ITGA5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RA   Morrisey E., Dutt P., Patel V.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 645-1053 FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=89235580; PubMed=2523953;
RA   Holers V.M., Ruff T.G., Parks D.L., McDonald J.A., Ballard L.L.,
RA   Brown E.J.;
RT   "Molecular cloning of a murine fibronectin receptor and its
RT   expression during inflammation. Expression of VLA-5 is increased in
RT   activated peritoneal macrophages in a manner discordant from major
RT   histocompatibility complex class II.";
RL   J. Exp. Med. 169:1589-1605(1989).
CC   -!- FUNCTION: INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRONECTIN
CC       AND FIBRINOGEN. IT RECOGNIZES THE SEQUENCE R-G-D IN ITS LIGANDS.
CC       MICE HOMOZYGOUS FOR A NULL MUTATION OF THE ALPHA-5 SUBUNIT GENE
CC       DIE AT DAY E10-E11. THEY SHOW BOTH EXTRA-EMBRYONIC AND EMBRYONIC
CC       VASCULAR DEFECTS, AND SEVERE ABNORMALITIES IN THE DEVELOPMENT OF
CC       THE POSTERIOR TRUNK. MAY PLAY A ROLE IN THE SURVIVAL OF ADULT
CC       SKELETAL MUSCLE.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC       SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC       DISULFIDE BOND. ALPHA-5 ASSOCIATES WITH BETA-1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC   -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79003; CAA55638.1; -.
DR   EMBL; X15203; CAA33273.1; -.
DR   PIR; S44250; S44250.
DR   HSSP; P06756; 1JV2.
DR   MGD; MGI:96604; Itga5.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 4.
DR   Pfam; PF00357; integrin_A; 1.
DR   SMART; SM00191; Int_alpha; 5.
DR   PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW   Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW   Signal; Calcium; Repeat.
FT   SIGNAL       1     44       BY SIMILARITY.
FT   CHAIN       45   1053       INTEGRIN ALPHA-5.
FT   CHAIN       45    898       INTEGRIN ALPHA-5 HEAVY CHAIN.
FT   CHAIN      899   1053       INTEGRIN ALPHA-5 LIGHT CHAIN.
FT   DOMAIN      45    999       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1000   1025       POTENTIAL.
FT   DOMAIN    1026   1053       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      60    122       FG-GAP 1.
FT   REPEAT     134    201       FG-GAP 2.
FT   REPEAT     202    259       FG-GAP 3.
FT   REPEAT     272    324       FG-GAP 4.
FT   REPEAT     326    390       FG-GAP 5.
FT   REPEAT     392    452       FG-GAP 6.
FT   REPEAT     456    508       FG-GAP 7.
FT   CA_BIND    337    345       POTENTIAL.
FT   CA_BIND    404    412       POTENTIAL.
FT   CA_BIND    468    476       POTENTIAL.
FT   SITE      1024   1028       GFFKR MOTIF.
FT   DISULFID   102    111       BY SIMILARITY.
FT   DISULFID   159    179       BY SIMILARITY.
FT   DISULFID   195    208       BY SIMILARITY.
FT   DISULFID   516    525       BY SIMILARITY.
FT   DISULFID   531    587       BY SIMILARITY.
FT   DISULFID   648    654       BY SIMILARITY.
FT   DISULFID   721    734       BY SIMILARITY.
FT   DISULFID   873    915       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   920    925       BY SIMILARITY.
FT   CARBOHYD    87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   185    185       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   300    300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   310    310       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   527    527       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   596    596       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   678    678       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   715    715       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   727    727       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   776    776       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   872    872       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1053 AA;  115056 MW;  DB872D11AC1755A6 CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 1053;
  Best Local Similarity   85.7%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              |||| ||
Db        288 DTEDFVA 294


Search completed: March 31, 2004, 16:46:17
Job time : 10.4667 secs
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE OF 1-10, CHARACTERIZATION, MASS SPECTROMETRY, AND
RP   MUTAGENESIS OF LYS-85.
RC   STRAIN=K12 / MC4100;
RX   MEDLINE=21179155; PubMed=11120740;
RA   Schuermann M., Sprenger G.A.;
RT   "Fructose-6-phosphate aldolase is a novel class I aldolase from
RT   Escherichia coli and is related to a novel group of bacterial
RT   transaldolases.";
RL   J. Biol. Chem. 276:11055-11061(2001).
CC   -!- CATALYTIC ACTIVITY: D-fructose 6-phosphate = glycerone + D-
CC       glyceraldehyde 3-phosphate.
CC   -!- ENZYME REGULATION: Inhibited by glycerol, inorganic phosphate and
CC       arabinose 5-phosphate.
CC   -!- SUBUNIT: Homodecamer or homododecamer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- MASS SPECTROMETRY: MW=22998; METHOD=Electrospray.
CC   -!- SIMILARITY: Belongs to the transaldolase family. Subfamily 3A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D88188; BAA13552.1; -.
DR   EMBL; AE000184; AAC73912.1; ALT_INIT.
DR   EMBL; D90720; BAA35513.1; ALT_INIT.
DR   EMBL; D90721; BAA35520.1; ALT_INIT.
DR   PDB; 1L6W; 12-JUN-02.
DR   EcoGene; EG13471; fsaA.
DR   HAMAP; MF_00496; -; 1.
DR   InterPro; IPR001585; Transaldolase.
DR   InterPro; IPR004731; Transaldolase_C.
DR   Pfam; PF00923; Transaldolase; 1.
DR   TIGRFAMs; TIGR00875; talC; 1.
DR   PROSITE; PS01054; TRANSALDOLASE_1; 1.
DR   PROSITE; PS00958; TRANSALDOLASE_2; 1.
KW   Lyase; Complete proteome; 3D-structure.
FT   ACT_SITE     85     85
FT   MUTAGEN      85     85       K->R: ALMOST COMPLETE LOSS OF ACTIVITY.
SQ   SEQUENCE   220 AA;  22997 MW;  317B9A06CD19C587 CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 220;
  Best Local Similarity   85.7%;  Pred. No. 40;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              || ||||
Db          6 DTSDVVA 12


RESULT 11
YWJI_BACSU
ID   YWJI_BACSU     STANDARD;      PRT;   321 AA.
AC   Q03224;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein ywjI.
GN   YWJI OR BSU37090.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BD99 / MS119;
RX   MEDLINE=93139031; PubMed=8423140;
RA   Quirk P.G., Dunkley E.A. Jr., Lee P., Krulwich T.A.;
RT   "Identification of a putative Bacillus subtilis rho gene.";
RL   J. Bacteriol. 175:647-654(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98015417; PubMed=9353933;
RA   Presecan E., Moszer I., Boursier L., Cruz Ramos H.C., De La Fuente V.,
RA   Hullo M.-F., Lelong C., Schleich S., Sekowska A., Song B.H.,
RA   Villani G., Kunst F., Danchin A., Glaser P.;
RT   "The Bacillus subtilis genome from gerBC (311 degrees) to licR (334
RT   degrees).";
RL   Microbiology 143:3313-3328(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: Belongs to the FBPase class II family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M97678; AAA02899.1; -.
DR   EMBL; Z49782; CAA89876.1; -.
DR   EMBL; Z99122; CAB15726.1; -.
DR   PIR; S55429; S55429.
DR   SubtiList; BG10415; ywjI.
DR   InterPro; IPR004464; GlpX.
```

```
DR   Pfam; PF03320; FBPase_glpX; 1.
DR   TIGRFAMs; TIGR00330; glpX; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   321 AA;  33951 MW;  418442320FCE7C09 CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 321;
  Best Local Similarity   85.7%;  Pred. No. 59;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              | |||||
Db        153 DVEDVVA 159


RESULT 12
TRME_STAEP
ID   TRME_STAEP     STANDARD;      PRT;   459 AA.
AC   Q8CMN5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   tRNA modification GTPase trmE.
GN   TRME OR SE2417.
OS   Staphylococcus epidermidis.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1282;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12228;
RX   PubMed=12950922;
RA   Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA   Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA   Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT   "Genome-based analysis of virulence genes in a non-biofilm-forming
RT   Staphylococcus epidermidis strain (ATCC 12228).";
RL   Mol. Microbiol. 49:1577-1593(2003).
CC   -!- FUNCTION: Exhibits a very high intrinsic GTPase hydrolysis rate.
CC       Involved in the biosynthesis of the hypermodified nucleoside 5-
CC       methylaminomethyl-2-thiouridine, which is found in the wobble
CC       position of some tRNAs (By similarity).
CC   -!- SIMILARITY: Belongs to the era/trmE family of GTP-binding
CC       proteins. TrmE subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016752; AAO06060.1; -.
DR   HAMAP; MF_00379; -; 1.
DR   InterPro; IPR005289; GTP-bindding_dom.
DR   InterPro; IPR002917; MMR_HSR1.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR004520; ThdF.
DR   Pfam; PF01926; MMR_HSR1; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   TIGRFAMs; TIGR00650; MG442; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   TIGRFAMs; TIGR00450; thdF; 1.
KW   tRNA processing; GTP-binding; Complete proteome.
FT   NP_BIND     228    235       GTP (POTENTIAL).
FT   NP_BIND     275    279       GTP (POTENTIAL).
FT   NP_BIND     335    338       GTP (POTENTIAL).
SQ   SEQUENCE   459 AA;  51451 MW;  BE425B02FB9D2AA1 CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 459;
  Best Local Similarity   83.3%;  Pred. No. 86;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||:|
Db        281 DTEDIV 286


RESULT 13
2A5R_MOUSE
ID   2A5R_MOUSE     STANDARD;      PRT;   491 AA.
AC   Q9Z176;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Protein phosphatase 2A, 59 kDa regulatory subunit B (PP2A PR59) (PP2A
DE   B''-PR59).
GN   PPP2R3A OR PPP2R6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Teratocarcinoma;
RX   MEDLINE=99124398; PubMed=9927208;
RA   Voorhoeve P.M., Hijmans E.M., Bernards R.;
RT   "Functional interaction between a novel protein phosphatase 2A
RT   regulatory subunit, PR59, and the retinoblastoma-related p107
RT   protein.";
RL   Oncogene 18:515-524(1999).
CC   -!- FUNCTION: The B regulatory subunit might modulate substrate
CC       selectivity and catalytic activity, and also might direct the
CC       localization of the catalytic enzyme to a particular subcellular
CC       compartment. Interacts with retinoblastoma-related protein p107
CC       (in vivo). May target PP2A core dimer to p107 resulting in
CC       dephosphorylation of p107.
CC   -!- SUBUNIT: PP2A consists of a common heterodimeric core enzyme,
CC       composed of a 36 kDa catalytic subunit (subunit C) and a 65 kDa
CC       constant regulatory subunit (PR65 or subunit A), that associates
CC       with a variety of regulatory subunits. Proteins that associate
CC       with the core dimer include three families of regulatory subunits
CC       B (the R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 and R5/B'/B56
CC       families), the 48 kDa variable regulatory subunit, viral proteins,
CC       and cell signaling molecules.
CC   -!- TISSUE SPECIFICITY: Expressed in testis, kidney, liver, lung,
CC       spleen, brain and heart.
CC   -!- SIMILARITY: Belongs to the phosphatase 2A regulatory subunit B
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF050165; AAC98973.1; -.
DR   MGD; MGI:1335093; Ppp2r3a.
DR   GO; GO:0008601; F:protein phosphatase type 2A, intrinsic regu. . .; IPI.
DR   GO; GO:0000080; P:G1 phase of mitotic cell cycle; IDA.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; IDA.
DR   GO; GO:0006470; P:protein amino acid dephosphorylation; IDA.
DR   InterPro; IPR002048; EF-hand.
DR   PROSITE; PS00018; EF_HAND; 1.
KW   Calcium-binding; Multigene family.
FT   CA_BIND     344    355       EF-HAND (POTENTIAL).
FT   DOMAIN      464    469       POLY-ASP.
SQ   SEQUENCE   491 AA;  55705 MW;  4B4BAB5C864FB683 CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 491;
  Best Local Similarity   71.4%;  Pred. No. 92;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
```

```
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   PDB; 1A1R; 17-JUN-98.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   TRANSFAC; T04155; -.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR002518; HCV_NS2.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_C29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   ProDom; PD186062; HCV_NS1; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   STRAND     1224   1226
FT   TURN       1232   1233
FT   TURN       1236   1238
FT   HELIX      1239   1246
FT   TURN       1247   1248
FT   STRAND     1251   1255
FT   HELIX      1258   1271
FT   TURN       1272   1272
FT   STRAND     1277   1280
FT   TURN       1281   1282
FT   STRAND     1283   1285
FT   STRAND     1291   1295
FT   HELIX      1296   1301
FT   TURN       1302   1303
FT   STRAND     1312   1316
FT   TURN       1317   1319
FT   HELIX      1323   1335
FT   TURN       1336   1340
FT   STRAND     1343   1347
FT   TURN       1352   1353
FT   TURN       1360   1361
FT   STRAND     1362   1366
FT   STRAND     1368   1368
```

```
FT   STRAND      1373   1375
FT   TURN        1376   1377
FT   STRAND      1378   1380
FT   HELIX       1382   1385
FT   STRAND      1389   1393
FT   HELIX       1397   1409
FT   TURN        1410   1411
FT   STRAND      1414   1417
FT   TURN        1419   1420
FT   STRAND      1432   1436
FT   TURN        1438   1439
FT   STRAND      1450   1453
FT   STRAND      1456   1463
FT   STRAND      1471   1478
FT   STRAND      1480   1480
FT   HELIX       1481   1488
FT   TURN        1489   1490
FT   STRAND      1497   1501
FT   STRAND      1507   1507
FT   STRAND      1511   1511
FT   HELIX       1514   1527
FT   HELIX       1532   1544
FT   STRAND      1550   1550
FT   HELIX       1555   1564
FT   HELIX       1570   1578
FT   TURN        1579   1580
FT   HELIX       1584   1597
FT   TURN        1598   1598
FT   HELIX       1606   1611
FT   TURN        1614   1618
FT   STRAND      1622   1623
FT   STRAND      1627   1627
FT   STRAND      1635   1636
FT   HELIX       1640   1652
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 3011;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db       2413 DTEDVV 2418


RESULT 9
FSAA_ECO57
ID   FSAA_ECO57     STANDARD;      PRT;   220 AA.
AC   P58423;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fructose-6-phosphate aldolase 1 (EC 4.1.2.-).
GN   FSAA OR FSA OR MIPB OR Z1048 OR ECS0903.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- CATALYTIC ACTIVITY: D-fructose 6-phosphate = glycerone + D-
CC       glyceraldehyde 3-phosphate.
CC   -!- SUBUNIT: Homodecamer or homododecamer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the transaldolase family. Subfamily 3A.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE005263; AAG55198.1; ALT_INIT.
DR   EMBL; AP002553; BAB34326.1; ALT_INIT.
DR   HAMAP; MF_00496; -; 1.
DR   InterPro; IPR001585; Transaldolase.
DR   InterPro; IPR004731; Transaldolase_C.
DR   Pfam; PF00923; Transaldolase; 1.
DR   TIGRFAMs; TIGR00875; talC; 1.
DR   PROSITE; PS01054; TRANSALDOLASE_1; 1.
DR   PROSITE; PS00958; TRANSALDOLASE_2; 1.
KW   Lyase; Complete proteome.
FT   ACT_SITE     85     85       BY SIMILARITY.
SQ   SEQUENCE   220 AA;  23113 MW;  B5D2B38FF13EDC5C CRC64;

  Query Match             82.9%;  Score 29;  DB 1;  Length 220;
  Best Local Similarity   85.7%;  Pred. No. 40;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              || ||||
Db          6 DTSDVVA 12


RESULT 10
FSAA_ECOLI
ID   FSAA_ECOLI     STANDARD;      PRT;   220 AA.
AC   P78055; P77855; Q9R3X3;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Fructose-6-phosphate aldolase 1 (EC 4.1.2.-).
GN   FSAA OR FSA OR MIPB OR B0825.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MC4100;
RA   Isomura M., Ogino T., Mizuno T.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
```

```
GN   ZP3.
OS   Macaca radiata (Bonnet monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9548;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96249321; PubMed=8848588;
RA   Kolluri S.K., Kaul R., Banerjee K., Gupta S.K.;
RT   "Nucleotide sequence of cDNA encoding bonnet monkey (Macaca radiata)
RT   zona pellucida glycoprotein-ZP3.";
RL   Reprod. Fertil. Dev. 7:1209-1212(1995).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82639; CAA57961.1; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    388    408       POTENTIAL.
FT   DOMAIN      409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    307       ZP.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  47040 MW;  3B41C4CFA3792331 CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 7
GYRA_HELPJ
ID   GYRA_HELPJ     STANDARD;      PRT;   828 AA.
AC   Q9ZLD9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA gyrase subunit A (EC 5.99.1.3).
GN   GYRA OR JHP0641.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: DNA gyrase negatively supercoils closed circular double-
CC       stranded DNA in an ATP-dependent manner and also catalyzes the
CC       interconversion of other topological isomers of double-stranded
CC       DNA rings, including catenanes and knotted rings.
CC   -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC       of double-stranded DNA.
CC   -!- SUBUNIT: Made up of two chains. The A chain is responsible for DNA
CC       breakage and rejoining; the B chain catalyzes ATP hydrolysis. The
CC       enzyme forms an A2B2 tetramer.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001496; AAD06219.1; -.
DR   PIR; E71906; E71906.
DR   HSSP; P09097; 1AB4.
DR   InterPro; IPR005743; DNA_gyrA.
DR   InterPro; IPR006691; DNA_gyraseA_C.
DR   InterPro; IPR002205; DNA_topoisoIV.
DR   Pfam; PF03989; DNA_gyraseA_C; 6.
DR   Pfam; PF00521; DNA_topoisoIV; 1.
DR   ProDom; PD000742; DNA_topoisoIV; 1.
DR   SMART; SM00434; TOP4c; 1.
DR   TIGRFAMs; TIGR01063; gyrA; 1.
KW   Topoisomerase; Isomerase; DNA-binding; Complete proteome.
FT   ACT_SITE    126    126       DNA CLEAVAGE (BY SIMILARITY).
SQ   SEQUENCE   828 AA;  92651 MW;  DE6E7DE4A7FBF150 CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 828;
  Best Local Similarity   71.4%;  Pred. No. 96;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVVA 7
              ||||::|
Db        497 DTEDLIA 503


RESULT 8
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11108;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
CC       WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S71825; AAB31866.1; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    388    408       POTENTIAL.
FT   DOMAIN      409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    307       ZP.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  46809 MW;  1DACBD03026C2739 CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 5
ZP3_HUMAN
ID   ZP3_HUMAN      STANDARD;      PRT;   424 AA.
AC   P21754; Q06633;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor)
DE   (ZP3A/ZP3B).
GN   ZP3 OR ZP3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ZP3A).
RX   MEDLINE=90349545; PubMed=2385582;
RA   Chamberlin M.E., Dean J.;
RT   "Human homolog of the mouse sperm receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6014-6018(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS ZP3A AND ZP3B).
RC   TISSUE=Ovary;
RX   MEDLINE=93122771; PubMed=1478648;
RA   van Duin M., Polman J.E., Verkoelen C.C., Bunschoten H.,
RA   Meyerink J.H., Olijve W., Aitken R.J.;
RT   "Cloning and characterization of the human sperm receptor ligand ZP3:
RT   evidence for a second polymorphic allele with a different frequency
RT   in the Caucasian and Japanese populations.";
RL   Genomics 14:1064-1070(1992).
CC   -!- FUNCTION: Functions as a sperm-receptor. It is responsible for
CC       sperm-adhesion to the zona pellucida, and may contribute to the
CC       species-specificity of the insemination.
CC   -!- SUBUNIT: ZP3 FORMS WITH ZP1 AND ZP2 THE ZONA PELLUCIDA, IN WHICH
CC       ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular
CC       matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=ZP3A;
CC         IsoId=P21754-1; Sequence=Displayed;
CC       Name=ZP3B;
CC         IsoId=P21754-2; Sequence=VSP_006949, VSP_006950;
CC   -!- PTM: Sulfated glycoprotein with O-linked oligosaccharides.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60504; AAA61336.1; -.
DR   EMBL; X56777; CAA40095.1; -.
DR   EMBL; A18567; CAA01398.1; -.
DR   PIR; A36000; A36000.
DR   Genew; HGNC:13189; ZP3.
DR   MIM; 182889; -.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
KW   Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW   Extracellular matrix; Multigene family; Alternative splicing.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    424       ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT   DOMAIN       23    387       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    388    408       POTENTIAL.
FT   DOMAIN      409    424       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       45    307       ZP.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    272    272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    364    372       PLIFLDRRG -> ATDLPGQEW (in isoform ZP3B).
FT                                /FTId=VSP_006949.
FT   VARSPLIC    373    424       Missing (in isoform ZP3B).
FT                                /FTId=VSP_006950.
FT   CONFLICT    345    345       R -> T (IN REF. 2).
SQ   SEQUENCE   424 AA;  47028 MW;  94517B7B66EFFE06 CRC64;

  Query Match             85.7%;  Score 30;  DB 1;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DTEDVV 6
              ||||||
Db         85 DTEDVV 90


RESULT 6
ZP3_MACRA
ID   ZP3_MACRA      STANDARD;      PRT;   424 AA.
AC   P53785;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE   glycoprotein ZP3) (Zona pellucida protein C) (Sperm receptor).
```